DEX-0315

BR

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2002 (22.08.2002)

PCT

(10) International Publication Number
WO 02/064788 A2

(51) International Patent Classification[7]: C12N 15/12, C07K 14/47, C12Q 1/68, C12N 5/10, C07K 16/18, G01N 33/53

(21) International Application Number: PCT/US01/45080

(22) International Filing Date:
20 November 2001 (20.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/252,054 20 November 2000 (20.11.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/252,054 (CIP)
Filed on 20 November 2000 (20.11.2000)

(71) Applicant *(for all designated States except US)*: **DIADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MACINA, Roberto, A.** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **RECIPON, Herve** [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). **CHEN, Sei-Yu** [—/US]; 160 Mira Street, Foster City, CA 94404 (US). **SUN, Yongming** [CN/US]; 551 Shoal Drive. Redwood City, CA 94065 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US).

(74) Agents: **LICATA, Jane, Massey** et al.; Licata & Tyrrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/064788 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO LUNG SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic lung cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating lung cancer and non-cancerous disease states in lung, identifying lung tissue, monitoring and identifying and/or designing agonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered lung tissue for treatment and research.

COMPOSITIONS AND METHODS RELATING TO LUNG SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application Serial No. 60/252,054 filed November 20, 2000, which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic lung cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating lung cancer and non-cancerous disease states in lung, identifying lung tissue and monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered lung tissue for treatment and research.

BACKGROUND OF THE INVENTION

Throughout the last hundred years, the incidence of lung cancer has steadily increased, so much so that now in many countries, it is the most common cancer. In fact, lung cancer is the second most prevalent type of cancer for both men and women in the United States and is the most common cause of cancer death in both sexes. Lung cancer deaths have increased ten-fold in both men and women since 1930, primarily due to an increase in cigarette smoking, but also due to an increased exposure to arsenic, asbestos, chromates, chloromethyl ethers, nickel, polycyclic aromatic hydrocarbons and other agents. *See* Scott, Lung Cancer: A Guide to Diagnosis and Treatment, Addicus Books (2000) and Alberg *et al.*, in Kane *et al.* (eds.) Biology of Lung Cancer, pp. 11-52, Marcel Dekker, Inc. (1998). Lung cancer may result from a primary tumor originating in the

lung or a secondary tumor which has spread from another organ such as the bowel or breast. Although there are over a dozen types of lung cancer, over 90% fall into two categories: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). *See* Scott, *supra*. About 20-25% of all lung cancers are characterized as SCLC, while 70-80% are diagnosed as NSCLC. *Id.* A rare type of lung cancer is mesothelioma, which is generally caused by exposure to asbestos, and which affects the pleura of the lung. Lung cancer is usually diagnosed or screened for by chest x-ray, CAT scans, PET scans, or by sputum cytology. A diagnosis of lung cancer is usually confirmed by biopsy of the tissue. *Id.*

SCLC tumors are highly metastatic and grow quickly. By the time a patient has been diagnosed with SCLC, the cancer has usually already spread to other parts of the body, including lymph nodes, adrenals, liver, bone, brain and bone marrow. *See* Scott, *supra*; Van Houtte *et al.* (eds.), Progress and Perspective in the Treatment of Lung Cancer, Springer-Verlag (1999). Because the disease has usually spread to such an extent that surgery is not an option, the current treatment of choice is chemotherapy plus chest irradiation. *See* Van Houtte, *supra*. The stage of disease is a principal predictor of long-term survival. Less than 5% of patients with extensive disease that has spread beyond one lung and surrounding lymph nodes, live longer than two years. *Id.* However, the probability of five-year survival is three to four times higher if the disease is diagnosed and treated when it is still in a limited stage, i.e., not having spread beyond one lung. *Id.*

NSCLC is generally divided into three types: squamous cell carcinoma, adenocarcinoma and large cell carcinoma. Both squamous cell cancer and adenocarcinoma develop from the cells that line the airways; however, adenocarcinoma develops from the goblet cells that produce mucus. Large cell lung cancer has been thus named because the cells look large and rounded when viewed microscopically, and generally are considered relatively undifferentiated. *See* Yesner, Atlas of Lung Cancer, Lippincott-Raven (1998).

Secondary lung cancer is a cancer initiated elsewhere in the body that has spread to the lungs. Cancers that metastasize to the lung include, but are not limited to, breast cancer, melanoma, colon cancer and Hodgkin's lymphoma. Treatment for secondary lung cancer may depend upon the source of the original cancer. In other words, a lung

cancer that originated from breast cancer may be more responsive to breast cancer treatments and a lung cancer that originated from the colon cancer may be more responsive to colon cancer treatments.

The stage of a cancer indicates how far it has spread and is an important indicator of the prognosis. In addition, staging is important because treatment is often decided according to the stage of a cancer. SCLC is divided into two stages: limited disease, *i.e.*, cancer that can only be seen in one lung and in nearby lymph nodes; and extensive disease, *i.e.*, cancer that has spread outside the lung to the chest or to other parts of the body. For most patients with SCLC, the disease has already progressed to lymph nodes or elsewhere in the body at the time of diagnosis. *See* Scott, *supra*. Even if spreading is not apparent on the scans, it is likely that some cancer cells may have spread away and traveled through the bloodstream or lymph system. In general, chemotherapy with or without radiotherapy is often the preferred treatment. The initial scans and tests done at first will be used later to see how well a patient is responding to treatment.

In contrast, non-small cell cancer may be divided into four stages. Stage I is highly localized cancer with no cancer in the lymph nodes. Stage II cancer has spread to the lymph nodes at the top of the affected lung. Stage III cancer has spread near to where the cancer started. This can be to the chest wall, the covering of the lung (pleura), the middle of the chest (mediastinum) or other lymph nodes. Stage IV cancer has spread to another part of the body. Stage I-III cancer is usually treated with surgery, with or without chemotherapy. Stage IV cancer is usually treated with chemotherapy and/or palliative care.

A number of chromosomal and genetic abnormalities have been observed in lung cancer. In NSCLC, chromosomal aberrations have been described on 3p, 9p, 11p, 15p and 17p, and chromosomal deletions have been seen on chromosomes 7, 11, 13 and 19. *See* Skarin (ed.), Multimodality Treatment of Lung Cancer, Marcel Dekker, Inc. (2000); Gemmill *et al.*, pp. 465-502, in Kane, *supra*; Bailey-Wilson *et al.*, pp. 53-98, in Kane, *supra*. Chromosomal abnormalities have been described on 1p, 3p, 5q, 6q, 8q, 13q and 17p in SCLC. *Id.* The loss of the short arm of chromosome 3p has also been seen in greater than 90% of SCLC tumors and approximately 50% of NSCLC tumors. *Id.*

A number of oncogenes and tumor suppressor genes have been implicated in lung cancer. *See* Mabry, pp. 391-412, in Kane, *supra* and Sclafani *et al.*, pp. 295-316, in

Kane, *supra*. In both SCLC and NSCLC, the p53 tumor suppressor gene is mutated in over 50% of lung cancers. *See* Yesner, *supra*. Another tumor suppressor gene, FHIT, which is found on chromosome 3p, is mutated by tobacco smoke. *Id.*; Skarin, *supra*. In addition, more than 95% of SCLCs and approximately 20-60% of NSCLCs have an absent or abnormal retinoblastoma (Rb) protein, another tumor suppressor gene. The *ras* oncogene (particularly K-*ras*) is mutated in 20-30% of NSCLC specimens and the c-*erbB2* oncogene is expressed in 18% of stage 2 NSCLC and 60% of stage 4 NSCLC specimens. *See* Van Houtte, *supra*. Other tumor suppressor genes that are found in a region of chromosome 9, specifically in the region of 9p21, are deleted in many cancer cells, including $p16^{INK4A}$ and $p15^{INK4B}$. *See* Bailey-Wilson, *supra*; Sclafani *et al.*, *supra*. These tumor suppressor genes may also be implicated in lung cancer pathogenesis.

In addition, many lung cancer cells produce growth factors that may act in an autocrine fashion on lung cancer cells. *See* Siegfried *et al.*, pp. 317-336, in Kane, *supra*; Moody, pp. 337-370, in Kane, *supra* and Heasley *et al.*, 371-390, in Kane, *supra*. In SCLC, many tumor cells produce gastrin-releasing peptide (GRP), which is a proliferative growth factor for these cells. *See* Skarin, *supra*. Many NSCLC tumors express epidermal growth factor (EGF) receptors, allowing NSCLC cells to proliferate in response to EGF. Insulin-like growth factor (IGF-I) is elevated in greater than 95% of SCLC and greater than 80% of NSCLC tumors; it is thought to function as an autocrine growth factor. *Id.* Finally, stem cell factor (SCF, also known as steel factor or kit ligand) and c-Kit (a proto-oncoprotein tyrosine kinase receptor for SCF) are both expressed at high levels in SCLC, and thus may form an autocrine loop that increases proliferation. *Id.*

Although the majority of lung cancer cases are attributable to cigarette smoking, most smokers do not develop lung cancer. Epidemiological evidence has suggested that susceptibility to lung cancer may be inherited in a Mendelian fashion, and thus have an inherited genetic component. Bailey-Wilson, *supra*. Thus, it is thought that certain allelic variants at some genetic loci may affect susceptibility to lung cancer. *Id.* One way to identify which allelic variants are likely to be involved in lung cancer susceptibility, as well as susceptibility to other diseases, is to look at allelic variants of genes that are highly expressed in lung.

The lung is susceptible to a number of other debilitating diseases as well, including, without limitation, emphysema, pneumonia, cystic fibrosis and asthma. *See* Stockley (ed.), Molecular Biology of the Lung, Volume I: Emphysema and Infection, Birkhauser Verlag (1999), hereafter Stockley I, and Stockley (ed.), Molecular Biology of the Lung, Volume II: Asthma and Cancer, Birkhauser Verlag (1999), hereafter Stockley II. The cause of many these disorders is still not well understood and there are few, if any, good treatment options for many of these noncancerous lung disorders. Thus, there also remains a need for understanding of various noncancerous lung disorders and for identify treatments for these diseases.

The development and differentiation of the lung tissue during embryonic development is also very important. All of the epithelial cells of the respiratory tract, including those of the lung and bronchi, are derived from the primitive endodermal cells that line the embryonic outpouching. *See* Yesner, *supra*. During embryonic development, multipotent endodermal stem cells differentiate into many different types of specialized cells, which include ciliated cells for moving inhaled particles, goblet cells for producing mucus, Kulchitsky's cells for endocrine function, and Clara cells and type II pneumocytes for secreting surfactant protein. *Id*. Improper development and differentiation may cause respiratory disorders and distress in infants, particularly in premature infants, whose lungs cannot produce sufficient surfactant when they are born. Further, some lung cancer cells, particularly small cell carcinomas, appear multipotent, and can spontaneously differentiate into a number of cell types, including small cell carcinoma, adenocarcinoma and squamous cell carcinoma. *Id*. Thus, a better understanding of lung development and differentiation may help facilitate understanding of lung cancer initiation and progression.

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop lung cancer, for diagnosing lung cancer, for monitoring the progression of the disease, for staging the lung cancer, for determining whether the lung cancer has metastasized and for imaging the lung cancer. There is also a need for better treatment of lung cancer. There is also a great need for diagnosing and treating noncancerous lung disorders such as emphysema, pneumonia, lung infection, pulmonary fibrosis, cystic fibrosis and asthma. There is also a need for compositions and methods of using compositions that are capable of identifying lung tissue for forensic

purposes and for determining whether a particular cell or tissue exhibits lung-specific characteristics.

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat lung cancer and non-cancerous disease states in lung; identify and monitor lung tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered lung tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to lung cells, lung tissue and/or the lung organ. These lung specific nucleic acids (LSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the LSNA is genomic DNA, then the LSNA is a lung specific gene (LSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to lung. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 116 through 208. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 115. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding an LSP, or that selectively hybridize or exhibit substantial sequence similarity to an LSNA, as well as allelic variants of a nucleic acid molecule encoding an LSP, and allelic variants of an LSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes an LSP or that comprises a part of a nucleic acid sequence of an LSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of an LSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the

transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of an LSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic acid molecule encodes all or a fragment of an LSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of an LSNA.

Another object of the invention is to provided methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is an LSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of an LSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating lung cancer and non-cancerous disease states in lung. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring lung tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered lung tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat lung cancer and non-cancerous disease states in lung. The invention provides methods of using the polypeptides of the invention to identify and/or monitor lung tissue, and to produce engineered lung tissue.

The agonists and antagonists of the instant invention may be used to treat lung cancer and non-cancerous disease states in lung and to produce engineered lung tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and

pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround

the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid

molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs

typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phoshoraniladate, phosphoroamidate, and the like. *See e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology"

interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence

hybridizes to a perfectly matched probe. *See* Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$T_m = 81.5°C + 16.6 (\log_{10}[Na^+]) + 0.41$ (fraction G + C) – 0.63 (% formamide) - (600/l)

where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5 (\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.35 (% formamide) - (820/l).

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5(\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.50 (% formamide) - (820/l).

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify

nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula: $T_m = 81.5°C + 16.6(\log_{10}[Na^+]) + 0.41(\text{fraction G+C}) - (600/N)$, wherein N is change length and the $[Na^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAS. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra.*

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one

exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding an LSP or is an LSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g.*, Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g.*, Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g.*, Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455

(1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable

of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises an LSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation,

covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H, ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$--, and –CH_2SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al., Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino

acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See* Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed.,

Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as -, -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, -N,N,N-trimethyllysine, -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino

acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

1) Serine (S), Threonine (T);
2) Aspartic Acid (D), Glutamic Acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g.*, Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)
Cost to extend a gap:	1 (default
Max. alignments:	100 (default)
Word size:	11 (default)
No. of descriptions:	100 (default)
Penalty Matrix:	BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', $F(ab')_2$, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an $F(ab')_2$ fragment is a bivalent fragment comprising two Fab

fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g.*, Ward *et al.*, *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g.*, Bird *et al.*, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g.*, Holliger *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al.*, *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany

it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than1 μM, preferably less than100 nM and most preferably less than 10 nM.

The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "lung specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the lung as compared to other tissues in the body. In a preferred embodiment, a "lung specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "lung specific" nucleic acid molecule or polypeptide is expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the lung or to lung cells or tissue or that are derived from such nucleic acid molecules. These isolated lung specific nucleic acids (LSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to lung, a lung-specific polypeptide (LSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 116 through 208. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through115.

An LSNA may be derived from a human or from another animal. In a preferred embodiment, the LSNA is derived from a human or other mammal. In a more preferred embodiment, the LSNA is derived from a human or other primate. In an even more preferred embodiment, the LSNA is derived from a human.

By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding an LSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode an LSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes an LSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 116 through 208. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 115.

In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an LSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an LSP under moderate stringency conditions. In a more preferred

embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an LSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 116 through 208. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 115. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding an LSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human LSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 116 through 208. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding an LSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 116 through 208, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding an LSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding an LSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to an LSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 115. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity

with an LSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 115, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with an LSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with an LSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one that exhibits sequence identity over its entire length to an LSNA or to a nucleic acid molecule encoding an LSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the LSNA or the nucleic acid molecule encoding an LSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 116 through 208 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 115. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the LSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is

experimentally produced by directed mutation of an LSNA. Further, the substantially similar nucleic acid molecule may or may not be an LSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is an LSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of an LSNA or a nucleic acid encoding an LSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes an LSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes an LSP comprising an amino acid sequence of SEQ ID NO: 116 through 208. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is an LSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 115. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is an LSP. However, in a preferred embodiment, the part encodes an LSP. In one aspect, the invention comprises a part of an LSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to an LSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of an LSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes an LSP. A part comprises at least 10 nucleotides, more preferably at least 15,

17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with ^{33}P, ^{32}P, and ^{35}S, such as -^{32}P-dATP, -^{32}P-dCTP, -^{32}P-dGTP, -^{32}P-dTTP, -^{32}P-3'dATP, -^{32}P-ATP, -^{32}P-CTP, -^{32}P-GTP, -^{32}P-UTP, -^{35}S-dATP, α-^{35}S-GTP, α-^{33}P-dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. *See* Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3'

hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see* Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*; *BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g.*, Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See* Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.),

Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties. Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. *See* Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S.

Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The Tm of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the Tm of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the Tm by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the Tm by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases

do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.*, Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably,

detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of an LSNA, such as deletions, insertions, translocations, and duplications of the LSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.*, Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify LSNA in, and isolate LSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A^+- selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.*, Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to LSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding an LSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 116 through 208. In another preferred embodiment, the probe or primer is derived from an LSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular

Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention

can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides

encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their

derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λGT10 and λGT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as ura3-52, his3-D1, leu2-D1, trp1-D1 and lys2-201.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda* , *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-

transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of

the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast _-mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early

gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the LSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β-globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PltetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to

be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverley, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that

carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from

those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g,* Cormack *et al.*, *Gene* 173: 33–38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g,*. Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.*, Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.*, Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See* also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable

expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation,

and acylation, and it is an aspect of the present invention to provide LSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database http://www.abrf.org/ABRF/Research Committees/deltamass/deltamass.html (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. Nucleic Acids Res. 29; 332-335 (2001) and http://www.glycosuite.com/ (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. Nucleic Acids Research, 27: 370-372 (1999) and http://www.cbs.dtu.dk/databases/OGLYCBASE/ (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. Nucleic Acids Res 27(1):237-239 (1999) and http://www.cbs.dtu.dk/databases/PhosphoBase/ (accessed October 19, 2001); or http://pir.georgetown.edu/pirwww/search/textresid.html (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur

in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g, p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the

desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors

typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (Sf9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae, Schizosaccharomyces pombe, Pichia pastoris, Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda, e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (*e.g.*, from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and

readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from lung are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human lung cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. *See, e.g.*, Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.*, Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.*, with $CaCl_2$, or a solution of Mg^{2+}, Mn^{2+}, Ca^{2+}, Rb^{+} or K^{+}, dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (*e.g.*, Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols

(BioRad, Richmond, CA, USA) (http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+}. Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with $CaPO_4$ or introduced using liposomal and noniiposomal lipid-based agents. Commercial kits are available for $CaPO_4$ transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent,

FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.*, Cheng *et al., Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al., Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.*, Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides

Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a lung specific polypeptide (LSP). In an even more preferred embodiment, the polypeptide

is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 116 through 208. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of an LSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 116 through 208. A polypeptide that comprises only a fragment of an entire LSP may or may not be a polypeptide that is also an LSP. For instance, a full-length polypeptide may be lung-specific, while a fragment thereof may be found in other tissues as well as in lung. A polypeptide that is not an LSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-LSP antibodies. However, in a preferred embodiment, the part or fragment is an LSP. Methods of determining whether a polypeptide is an LSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g.*, Lerner, Nature 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping) and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, an LSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See*, *e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment of polypeptide of the invention, preferably an LSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably an LSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants, fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence

of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be lung-specific. In a preferred embodiment, the mutein is lung-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 116 through 208. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an LSP comprising an amino acid sequence of SEQ ID NO: 116 through 208. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to an LSP comprising an amino acid sequence of SEQ ID NO: 116 through 208.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is lung-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*,

Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the polypeptide is homologous to an LSP. In an even more preferred embodiment, the polypeptide is homologous to an LSP selected from the group having an amino acid sequence of SEQ ID NO: 116 through 208. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to an LSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ ID NO: 116 through 208. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an LSP comprising an amino acid sequence of SEQ ID NO: 116 through 208. In a yet more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to an LSP comprising an amino acid sequence of SEQ ID NO: 116 through 208. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to an LSP comprising an amino acid sequence of SEQ ID NO: 116 through 208. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to an LSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to an LSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the LSNA is selected from the group consisting of SEQ ID NO: 1 through 115. In another preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes an LSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred

embodiment, the LSP is selected from the group consisting of SEQ ID NO: 116 through 208.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus monkey, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 116 through 208. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the LSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of an LSP. Further, the homologous protein may or may not encode polypeptide that is an LSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is an LSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding an LSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 116 through 208. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 115.

In another embodiment, the invention provides polypeptides which comprise derivatives of the polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is an LSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 116 through 208, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H. In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed. W.H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslational events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive and thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, are suited for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514,

Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, e.g., APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSC, DSG (Pierce's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAMA, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyrogiobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-LSP antibodies.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of peptides administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Therap. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 413-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999), incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated

with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is an LSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 116 through 206. In a preferred embodiment, the analog is one that comprises one or more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to an LSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --CH(OH)CH_2-- and –CH_2SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of an LSP with the same amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g.*, Kole *et al.*, *Biochem. Biophys. Res. Com.* 209: 917-921 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (The Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (March 2000); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino ... analogues having detectable labels are also usefully incorporated during synthes... ide derivatives and analogs. Biotin, for example can be added using biotinoyl ... ...ymieuno-carbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugen... ...A... Biotin can also be added enzymatically by incorporation into a fusion p... ... *E. coli* B...A substrate peptide. The FMOC and *t*BOC derivatives of ...-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the ...dcyl chromophore at selected sites in the peptide sequence during synthesis. The ...phthalene derivative EDANS, the most common fluorophore for pairing with th... ...quencher ... fluorescence resonance energy transfer (FRET) systems, can b... ...ced during automated synthesis of peptides by using EDANS-FMOC ... acid ... the corresponding *t*BOC derivative (both from Molecular Pro... ..., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated d... automated FMOC synthesis of peptides using (FMOC)-TMR-... lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other ... amino acid analogues that can be incorporated during chemical synthesis inclu... ...tic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain pr... ... Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits ... ...s of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylat... ...ues.

A large ... of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo ...heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2... ...-endo-... ...oxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-... ... acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amin... ...clohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-c... ...-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanec...boxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-aminobenz... ... (antranilic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzo... ... Fmoc-2-am... ...benzophenone-2'-carboxylic acid, Fmoc-N-(4-

aminobenzoyl[illegible]ne, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric [illegible]moc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic [illegible]moc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic [illegible]moc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic [illegible]moc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic [illegible]moc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic [illegible]moc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic [illegible]moc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic [illegible]moc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropion[illegible]moc-L-methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxy[illegible]moc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from Advanced ChemTech Peptide Laboratory (Richmond, CA, USA).

Non-natural amino acid residues can also be added biosynthetically by engineering a suppressor tRNA that typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu et al., *Proc. Natl. Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5[illegible]):498-500 (2001).

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is an LSP. In a more preferred embodiment, the polypeptide unit fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 116 through 208, or is a mutein, homologous polypeptide, [illegible] derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic

acid sequence of SEQ ID NO: 1 through 115, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 115.

The fusion proteins of the present invention will include at least one fragment of the protein of the invention, which fragment is at least 6, typically at least 8, often at least 15, and usually at least 16, 18, 19, or 20 amino acids long. The fragment of the protein of the invention to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 25 amino acids in length. Fusions that include larger polypeptides, such as the GST (glutathione-S-transferase) fusion, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinant protein, *e.g.*, Ausubel, Chapter 16, (1992), *supra*. Although purification tags can be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification is not required; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also described above, heterologous polypeptides to be included in the fusion proteins of the present invention usefully include those that facilitate secretion of recombinant protein – into the periplasmic space or extracellular milieu for prokaryotic cells, or into the culture medium for eukaryotic cells — through incorporation of secretion signals and/or leader sequences. For example, a His^6 tagged protein can be purified on a nickel affinity column and a GST fusion protein can be purified on a

glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified [illegible] Protein A or Protein G column and a fusion protein comprising an epitope tag such [illegible] can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferred that the epitope tag be separated from the protein encoded by the [illegible] gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. *See* Bartel *et al.* (eds.), The [illegible]-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid [illegible]gies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); [illegible] *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin.* [illegible] 6([illegible]) 94 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17([illegible]) 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2 *Nature* 380: 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological [illegible] ays. *Science* 285, 591-595, Fabbrizio *et al.*, (1999) Inhibition of [illegible] cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register [illegible] relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nucleic Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in Drosophila with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; [illegible] *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial [illegible] *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; [illegible] M. (2000) A comprehensive analysis of protein-protein interactions in [illegible]omyces cerevisiae. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast

GAL4 DNA [illegible] bait plasmids are available that express the bait fused to a nuclear localization sequence.

Other useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated herein by reference in its entirety.

The proteins and fragments of the present invention can also usefully be fused to proteins such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin, [illegible] factor, and the like in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion proteins include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulin, β-galactosidase, biotin trpE, protein A, β-lactamase, α-amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino or carboxy terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast transcription factor, transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Kroll (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins can typically be made by either recombinant nucleic acid methods, as described above, or chemically synthesized using techniques well-known in the art (*e.g.*, a Merrifield synthesis) or by chemical cross-linking.

Another type of fusion proteins is that the epitope tag can be used to bind the fusion protein to a surface through an affinity linkage for screening binding proteins or other molecules that bind to the LSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize LSPs, their allelic variants and fragments. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the presence of the present invention, particularly LSPs, *e.g.* by ELISA for detection of proteins in liquid samples such as serum, by immunohistochemistry or laser scanning cytometry for detection of protein in tissue samples, or by flow cytometry, for

detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of ESPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of ESPs.

One may determine which polypeptides including muteins, fusion proteins, homologous proteins and the like are functional by methods known in the art. For instance, residues tolerant to change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis (Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000) followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ <Tet> Linker™ Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies, Madison, WI, USA)

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. See, *e.g.*, Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

According to an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are those that lack detectable ampholytes, acrylamide monomers, bisacrylamide and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can be attached to a substrate. The substrate can be porous or solid, planar or non-planar, and the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, nylon membrane, and the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized proteins of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized proteins of the present invention. Such plastics include polymethylacrylate, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, cellulose acetate, cellulose nitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a microtiter well, the microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The polypeptides, fragments, analogs, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the polypeptide, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that comprises an LSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 116 through 208, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as expressed, e.g., following solubilization in SDS. New epitopes may be also due to a different post-translational modifications (PTMs) in disease versus normal tissue. For example, a different PTM on a LSP may be glycosylated in cancerous cells, but not glycosylated in normal cells, or visa versa. In addition, alternative splice forms of a LSP may be indicative of cancer. Differential degradation of the C or N-terminus of a LSP may also be a marker or target for anticancer therapy. For example, a LSP may be N-terminally degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is known in the art, the degree to which an antibody can discriminate as among molecules in a sample will depend, in part, upon the conformational relatedness of the epitopes in the sample; typically, the antibodies of the present invention will discriminate over adventitious binding to non-LSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the present invention in samples derived from human lung.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM antibody) of the present invention for a protein or protein fragment of the present invention will be at least about 1 x 10^{-6} molar (M), typically at least about 5 x 10^{-7}

^{7} M, 1 x 10^{-7} ... ...dities of at least 1 x 10^{-8} M, 5 x 10^{-9} M, 1 x 10^{-10} M and up to 1 X ... ...ny useful

The an... ...invention can be naturally-occurring forms, such as IgG, IgM, IgD ... ...in any avian, reptilian, or mammalian species.

Human ... ...can, ...l infrequently, be drawn directly from human donors or human ... ...in this ... ...antibodies to the proteins of the present invention will typically ... ...obtained from ... ...neous immunization, such as autoimmune immunization. ... ...the protein ... ...otein fragments of the present invention. Such antibodies wi... ... ...invariably be polyclonal. In addition, individual polyclonal an... ... ...and cloned to generate monoclonals.

Human ... ...quently obtained using transgenic animals that express huma... ...2108... ..., which transgenic animals can be affirmatively immunized wi... ..., protein immu... ...gen of the present invention. Human Ig-transgenic mice capable of ...producing huma... ...ntibodies and methods of producing human antibodies the... ...upon spec... ...immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,... ...14,39... ...75,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,... ...7,0... ...1,016; 5,6... ...25; 5,625,126; 5,569,825; 5,545,807; 5,... ...the disclosures of which are incorporated herein by reference in t... ...bodies are typically monoclonal, and are typically produced using ... ...for production of murine antibodies.

Human ... ...bodies are par... ...larly useful, and often preferred, when the antibodies of the present invention ... ...to be administered to human beings as *in vivo* diagnostic or t...peutic agents. ...e recipient immune response to the administered antibody will ... ...s than that occasioned by administration of an antibody deri... ... ...such a... ...se.

IgG, i... ... ...gA antibodies of the present invention can also be obtained from ... ...mammals such as rodents (typically mouse, but also rat, guine... ... ...hamster) ...morphs, typically rabbits, and also larger mammals, such ... ...sheep, goats, ... and horses, and other egg laying birds or reptiles such as chicke... ...or alligators. ... ...example, avian antibodies may be generated using techniques des...ribed in WO 00/2... ...4, published 25 May 2000, the contents of which are hereby incorpo... ...ir... ...n such cases, as with the transgenic human-

antibody-prod... ...mals, fortuitous immunization is not required, and the non-huma... ...firmatively immunized, according to standard immunization ... with the ...ein or protein fragment of the present invention.

As disc... ...bove, virtua... all fragments of 8 or more contiguous amino acids of the proteins of the present inve...ion can be used effectively as immunogens when conjugated to a ...rier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, o... ...serum albumin, conveniently using a bifunctional linker such as those describe... ...h discussion is incorporated by reference here.

Immun... ...ty can also ...conferred by fusion of the polypeptide and fragments of the present ... ...to othe... ...ties. For example, peptides of the present invention can be produce... ...id phase ...thesis on a branched polylysine core matrix; these multiple antigen... ...peptides (MAP...) provide high purity, increased avidity, accurate chemical defin...ion and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. US...* ...69-5413 ...'88); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protoc... ...-human mammals or avian species are well-established in ... ...Harlo... *al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring ...Laboratory ...'88); Coligan *et al.* (eds.), Current Protocols in Immunology, ...Wiley & Son... Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Mo...clonal Antibod...and Engineered Antibody Derivatives (Basics: From Background to ...ques), Springe... ...rlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* ...7-422 (1... ...the disclosures of which are incorporated herein by reference. In... ...ten include multiple immunizations, either with or without adjuv... ...as Fr... ...omplete adjuvant and Freund's incomplete adjuvant, and may incl... ...DNA ...unization (Moss, *Semin. Immunol.* 2: 317-327 (1990).

Antib... ...to all non-... ...mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the inven... ...ntibodies from avian species may have particular advantage in ... ...s of the present invention, in human serum or tissues (Vikinge et ... *Biotech...* 13: 1257-1262 (1998).

Following ... antibodies ... the present invention can be produced using any art-a... techniques are well-known in the art, Coligan, *supra*; Zola, *su...* ...ard ...), Basic Methods in Antibody Production and Characterization ... Press (200... Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. ... Humana Press ...995); Delves (ed.), Antibody Production: Essential Techniques, Jon... Wiley & Son ...1997); Kenney, Antibody Solution: An Antibody Methods Manu... ...ghan & P... ...1997), incorporated herein by reference in their entireties, and ... not ... d here.

Briefly, ... such ... ques include, *inter alia*, production of monoclonal antibodies by i... ... and ... ion of antibodies or fragments or derivatives thereof from i... ... engineered ... express immunoglobulin genes or fragments thereof. These ... methods of p... ...ction are not mutually exclusive: genes encoding antibodies specific for the protein... protein fragments of the present invention can be cloned from h... ...mas and the... ...er expressed in other host cells. Nor need the two necessarily be ... ...er together ..., genes encoding antibodies specific for the proteins and p... ...ment... ... present i... ...ation can be cloned directly from B cells known to ... ...e for th... ...ired protein, as further described in U.S Patent 5,627,052, th... ... of ... ... incorporated herein by reference in its entirety, or from antibody-... ...ng phage.

Recombinant expression ... host cells is particularly useful when fragments or derivatives of the antibodies of th... present invention are desired.

Host ce... for recombinant production of either whole antibodies, antibody fragments, or ... ...erivat... ...n be prokaryotic or eukaryotic.

Prokar... ... are p... ...arly useful for producing phage displayed antibodies of the present ...

The te... ...y of pha... ...played antibodies, in which antibody variable region fragments are ...ed, for example, ... the gene iii protein (pIII) or gene VIII protein (pVIII) for display on the surface ... filamentous phage, such as M13, is by now well-established. ... ... Siddiu... ... *Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Op...* ... ...2-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (... ... ...*ent Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et a...* ... ... 155-168 (1997); Hoogenboom, *Trends in*

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in*

Biotechnol. 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.*, Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.*, Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, J. Biotechnol. 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.*, Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et*

al., *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock et al., *J. Immunol Methods.* 231: 147-57 (1999); Young et al., *Res. Immunol.* 149: 609-10 (1998); Limonta et al., *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem.* (Tokyo) 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entireties.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, $F(ab)'_2$, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that

[redacted]

invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or

one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA*.81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or

transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β-galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopryanoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic

compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g.*, Thorpe *et al., Methods Enzymol.* 133: 331-53 (1986); Kricka *et al., J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al., J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P, ^{32}P, ^{35}S, ^{3}H, and ^{125}I.

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th, ^{227}Ac, ^{225}Ac, ^{223}Ra, ^{213}Bi, ^{212}Pb, ^{212}Bi, ^{211}At, ^{203}Pb, ^{194}Os, ^{188}Re, ^{186}Re, ^{153}Sm, ^{149}Tb, ^{131}I, ^{125}I, ^{111}In, ^{105}Rh, ^{99m}Tc, ^{97}Ru, ^{90}Y, ^{90}Sr, ^{88}Y, ^{72}Se, ^{67}Cu, or ^{47}Sc.

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. *See* Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding an LSP. In a preferred embodiment, the LSP comprises an amino

acid sequence selected from SEQ ID NO: 116 through 208, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise an LSNA of the invention, preferably an LSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 115, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human LSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.*, Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.*, Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989 retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.*, Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.*, Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.*, Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.*, Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.*, Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.*, Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e.*, a

nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g.*, the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of

the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g.*, Gu *et al.*, *Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g.*, Smithies *et al.*, *Nature* 317: 230-234 (1985); Thomas *et al.*, *Cell* 51: 503-512 (1987); Thompson *et al.*, *Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g.*, Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC

compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 115 and SEQ ID NO: 116 through 208 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set

representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Lung Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by

comparing expression of an LSNA or an LSP in a human patient that has or may have lung cancer, or who is at risk of developing lung cancer, with the expression of an LSNA or an LSP in a normal human control. For purposes of the present invention, "expression of an LSNA" or "LSNA expression" means the quantity of LSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of an LSP" or "LSP expression" means the amount of LSP that can be measured by any method known in the art or the level of translation of an LSG LSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing lung cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of LSNA or LSP in cells, tissues, organs or bodily fluids compared with levels of LSNA or LSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of an LSNA or LSP in the patient versus the normal human control is associated with the presence of lung cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing lung cancer in a patient by analyzing changes in the structure of the mRNA of an LSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing lung cancer in a patient by analyzing changes in an LSP compared to an LSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the LSP or subcellular LSP localization.

In a preferred embodiment, the expression of an LSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 116 through 208, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the LSNA expression that is measured is the level of expression of an LSNA mRNA selected from SEQ ID NO: 1 through 115, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. LSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by

Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. LSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of an LSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, LSNA expression may be compared to a known control, such as normal lung nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of an LSP is measured by determining the level of an LSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 116 through 208, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of LSNA or LSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of lung cancer. The expression level of an LSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the LSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See*, e.g, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the LSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to an LSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-LSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a

protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the LSP will bind to the anti-LSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-LSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the LSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of an LSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure LSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-LSP antibody is attached to a solid support and an allocated amount of a labeled LSP and a sample of interest are incubated with the solid support. The amount of labeled LSP detected which is attached to the solid support can be correlated to the quantity of an LSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of an LSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other

mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more LSNAs of interest. In this approach, all or a portion of one or more LSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of LSNA or LSP includes, without limitation, lung tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, lung cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary lung cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediatinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See* Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in LSNAs or LSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of an LSNA or LSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other LSNA or LSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular LSNA or LSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more LSNAs and/or LSPs in a sample from a patient suspected of having lung cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of an LSNA and/or LSP and then ascertaining whether the patient has lung cancer from the expression level of the LSNA or LSP. In general, if high expression relative to a control of an LSNA or LSP is indicative of lung cancer, a diagnostic assay is considered positive if the level of expression of the LSNA or LSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an LSNA or LSP is indicative of lung cancer, a diagnostic assay is considered positive if the level of expression of the LSNA or LSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether lung cancer has metastasized in a patient. One may identify whether the lung cancer has metastasized by measuring the expression levels and/or structural alterations of one or more LSNAs and/or LSPs in a variety of tissues. The presence of an LSNA or LSP in a certain tissue

at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of an LSNA or LSP is associated with lung cancer. Similarly, the presence of an LSNA or LSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of an LSNA or LSP is associated with lung cancer. Further, the presence of a structurally altered LSNA or LSP that is associated with lung cancer is also indicative of metastasis.

In general, if high expression relative to a control of an LSNA or LSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the LSNA or LSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an LSNA or LSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the LSNA or LSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The LSNA or LSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with lung cancers or other lung related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of lung disorders.

Staging

The invention also provides a method of staging lung cancer in a human patient. The method comprises identifying a human patient having lung cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more LSNAs or LSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more LSNAs or LSPs is determined for each stage to obtain a standard expression level for each LSNA and LSP. Then, the LSNA or LSP expression levels are determined in a biological sample from a patient whose stage of cancer is not known. The LSNA or LSP expression levels from the patient are then compared to the

standard expression level. By comparing the expression level of the LSNAs and LSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of an LSNA or LSP to determine the stage of a lung cancer.

Monitoring

Further provided is a method of monitoring lung cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the lung cancer. The method comprises identifying a human patient that one wants to monitor for lung cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more LSNAs or LSPs, and comparing the LSNA or LSP levels over time to those LSNA or LSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in an LSNA or LSP that are associated with lung cancer.

If increased expression of an LSNA or LSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of an LSNA or LSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of an LSNA or LSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of an LSNA or LSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of LSNAs or LSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of lung cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of an LSNA and/or LSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more LSNAs and/or LSPs are detected. The presence of higher (or lower) LSNA or LSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly lung cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more LSNAs and/or LSPs of the invention can also be monitored by analyzing levels of expression of the LSNAs and/or LSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in an LSG, thereby determining if a human with the genetic lesion is susceptible to developing lung cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing lung cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the LSGs of this invention, a chromosomal rearrangement of LSG, an aberrant modification of LSG (such as of the methylation pattern of the genomic DNA), or allelic loss of an LSG. Methods to detect such lesions in the LSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous Lung Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more LSNAs and/or LSPs in a sample from a patient suspected of having or known to have a noncancerous lung disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of an LSNA and/or LSP, comparing the expression level or structural alteration of the LSNA or LSP to a normal lung control,

and then ascertaining whether the patient has a noncancerous lung disease. In general, if high expression relative to a control of an LSNA or LSP is indicative of a particular noncancerous lung disease, a diagnostic assay is considered positive if the level of expression of the LSNA or LSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an LSNA or LSP is indicative of a noncancerous lung disease, a diagnostic assay is considered positive if the level of expression of the LSNA or LSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether an LSNA and/or LSP is associated with a particular noncancerous lung disease by obtaining lung tissue from a patient having a noncancerous lung disease of interest and determining which LSNAs and/or LSPs are expressed in the tissue at either a higher or a lower level than in normal lung tissue. In another embodiment, one may determine whether an LSNA or LSP exhibits structural alterations in a particular noncancerous lung disease state by obtaining lung tissue from a patient having a noncancerous lung disease of interest and determining the structural alterations in one or more LSNAs and/or LSPs relative to normal lung tissue.

Methods for Identifying Lung Tissue

In another aspect, the invention provides methods for identifying lung tissue. These methods are particularly useful in, *e.g.*, forensic science, lung cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is lung tissue or has lung tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising lung tissue or having lung tissue-like characteristics, determining whether the sample expresses one or more LSNAs and/or LSPs, and, if the sample expresses one or more LSNAs and/or LSPs, concluding that the sample comprises lung tissue. In a preferred embodiment, the LSNA encodes a

polypeptide having an amino acid sequence selected from SEQ ID NO: 116 through 208, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the LSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 115, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses an LSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether an LSP is expressed. Determining whether a sample expresses an LSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the LSP has an amino acid sequence selected from SEQ ID NO: 116 through 208, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two LSNAs and/or LSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five LSNAs and/or LSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is lung tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into lung tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new lung tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying Lung Tissue

In another aspect, the invention provides methods for producing engineered lung tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing an LSNA or an LSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of lung tissue cells. In a preferred

embodiment, the cells are pluripotent. As is well-known in the art, normal lung tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered lung tissue or cells comprises one of these cell types. In another embodiment, the engineered lung tissue or cells comprises more than one lung cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the lung cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more LSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode LSPs having amino acid sequences selected from SEQ ID NO: 116 through 208, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 115, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, an LSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial lung tissue may be used to treat patients who have lost some or all of their lung function.

Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises an LSNA or part thereof. In a more preferred embodiment, the LSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 115, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises an LSP or fragment thereof. In a more preferred embodiment, the LSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 116 through 208, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the

pharmaceutical composition comprises an anti-LSP antibody, preferably an antibody that specifically binds to an LSP having an amino acid that is selected from the group consisting of SEQ ID NO: 116 through 208, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot

injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts

tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example LSP polypeptide, fusion protein, or fragments thereof, antibodies specific for LSP, agonists, antagonists or inhibitors of LSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age,

weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention , *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of lung function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication

incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV) , for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of an LSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of an LSP are administered, for example, to complement a deficiency in the native LSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid molecule encodes an LSP having the amino acid sequence of SEQ ID NO: 116 through 208, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express an LSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in LSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode an LSP having the amino acid sequence of SEQ ID NO: 116 through 208, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of an LSG antisense nucleic acid, are administered to downregulate transcription and/or translation of an LSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of an LSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to LSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the LSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding an LSP, preferably an LSP comprising an amino acid sequence of SEQ ID NO: 116 through 208, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an LSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant LSP defect.

Protein compositions are administered, for example, to complement a deficiency in native LSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to LSP. The immune response can

be used to modulate activity of LSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate LSP.

In a preferred embodiment, the polypeptide is an LSP comprising an amino acid sequence of SEQ ID NO: 116 through 208, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of LSP, or to target therapeutic agents to sites of LSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to an LSP comprising an amino acid sequence of SEQ ID NO: 116 through 208, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to an LSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to an LSP or have a modulatory effect on the expression or activity of an LSP. Modulators which decrease the expression or activity of LSP (antagonists) are believed to be useful in treating lung cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of an LSP can also be designed, synthesized and tested for use in the imaging and treatment of lung cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the LSPs identified herein. Molecules identified in the library as being capable of binding to an LSP are key candidates for

further evaluation for use in the treatment of lung cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of an LSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of LSP is administered. Antagonists of LSP can be produced using methods generally known in the art. In particular, purified LSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of an LSP.

In other embodiments a pharmaceutical composition comprising an agonist of an LSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an LSP comprising an amino acid sequence of SEQ ID NO: 116 through 208, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an LSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Lung Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the lung or to specific cells in the lung. In a preferred embodiment, an anti-LSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if lung tissue needs to be selectively destroyed. This would be useful for targeting and killing lung cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting lung cell function.

In another embodiment, an anti-LSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring lung function, identifying lung cancer tumors, and identifying noncancerous lung diseases.

EXAMPLES

Example 1: Gene Expression analysis

LSGs were identified by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc (Palo Alto, CA) using the data mining software package CLASP™ (Candidate Lead Automatic Search Program). CLASP™ is a set of algorithms that interrogate Incyte's database to identify genes that are both specific to particular tissue types as well as differentially expressed in tissues from patients with cancer. LifeSeq® Gold contains information about which genes are expressed in various tissues in the body and about the dynamics of expression in both normal and diseased states. CLASP™ first sorts the LifeSeq® Gold database into defined tissue types, such as breast, ovary and prostate. CLASP™ categorizes each tissue sample by disease state. Disease states include "healthy," "cancer," "associated with cancer," "other disease" and "other." Categorizing the disease states improves our ability to identify tissue and cancer-specific molecular targets. CLASP™ then performs a simultaneous parallel search for genes that are expressed both (1) selectively in the defined tissue type compared to other tissue types and (2) differentially in the "cancer" disease state compared to the other disease states affecting the same, or different, tissues. This sorting is accomplished by using mathematical and statistical filters that specify the minimum change in expression levels and the minimum frequency that the differential expression pattern must be observed across the tissue samples for the gene to be considered statistically significant. The CLASP™ algorithm quantifies the relative abundance of a particular gene in each tissue type and in each disease state.

To find the LSGs of this invention, the following specific CLASP™ profiles were utilized: tissue-specific expression (CLASP 1), detectable expression only in cancer tissue (CLASP 2), highest differential expression for a given cancer (CLASP 4); differential expression in cancer tissue (CLASP 5), and. cDNA libraries were divided into 60 unique tissue types (early versions of LifeSeq® had 48 tissue types). Genes or ESTs were grouped into "gene bins," where each bin is a cluster of sequences grouped together where they share a common contig. The expression level for each gene bin was calculated for each tissue type. Differential expression significance was calculated with rigorous statistical significant testing taking into account variations in sample size and

relative gene abundance in different libraries and within each library (for the equations used to determine statistically significant expression see Audic and Claverie "The significance of digital gene expression profiles," Genome Res 7(10): 986-995 (1997), including Equation 1 on page 987 and Equation 2 on page 988, the contents of which are incorporated by reference). Differentially expressed tissue-specific genes were selected based on the percentage abundance level in the targeted tissue versus all the other tissues (tissue-specificity). The expression levels for each gene in libraries of normal tissues or non-tumor tissues from cancer patients were compared with the expression levels in tissue libraries associated with tumor or disease (cancer-specificity). The results were analyzed for statistical significance.

For some of the nucleotide sequences found by mRNA subtraction, the following tissue expression levels were observed:

DEX0273_18	SEQ ID NO: 18	BRN .001	KID .0013	THY .002	TST .0027
DEX0273_19	SEQ ID NO: 19	BRN .001	KID .0013	THY .002	TST .0027
DEX0273_39	SEQ ID NO: 39	LIV .0019			
DEX0273_40	SEQ ID NO: 40	LIV .0019			
DEX0273_66	SEQ ID NO: 66	SAG .1383	PIT .2301	BMR .2381	URE .2474
DEX0273_69	SEQ ID NO: 69	SAG .1383	PIT .2301	BMR .2381	URE .2474
DEX0273_70	SEQ ID NO: 70	SAG .1383	PIT .2301	BMR .2381	URE .2474
DEX0273_88	SEQ ID NO: 88	SAG .1383	PIT .2301	BMR .2381	URE .2474

Abbreviation for tissues:
BLO Blood; BRN Brain; CON Connective Tissue; CRD Heart; FTS Fetus; INL Intestine, Large; INS Intestine, Small; KID Kidney; LIV Liver; LNG Lung; MAM Breast; MSL Muscles; NRV Nervous Tissue; OVR Ovary; PRO Prostate; STO Stomach; THR Thyroid Gland; TNS Tonsil / Adenoids; UTR Uterus

The chromosomal locations for the sequences are as follows:

DEX0273_1 chromosome 4
DEX0273_3 chromosome 1
DEX0273_4 chromosome 22
DEX0273_8 chromosome 9
DEX0273_9 chromosome 9
DEX0273_31 chromosome 20
DEX0273_32 chromosome 16
DEX0273_33 chromosome 16
DEX0273_35 chromosome 9
DEX0273_40 chromosome 10
DEX0273_41 chromosome 9
DEX0273_42 chromosome 9
DEX0273_48 chromosome 6
DEX0273_56 chromosome 22

DEX0273_59 chromosome 3
DEX0273_60 chromosome 10
DEX0273_64 chromosome 1
DEX0273_66 chromosome 8
DEX0273_67 chromosome 8
DEX0273_70 chromosome 8
DEX0273_71 chromosome 17
DEX0273_81 chromosome 12
DEX0273_89 chromosome 8
DEX0273_97 chromosome 22
DEX0273_103 chromosome 19
DEX0273_106 chromosome 21
DEX0273_108 chromosome 22
DEX0273_111 chromosome 9
DEX0273_112 chromosome 6

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'- 3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The

results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the LSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal tissue (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the LSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal tissue (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, the LSNAs show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 115 being diagnostic markers for cancer.

Example 3: Protein Expression

The LSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the LSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the LSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH_2-terminus of the coding sequence of LSNA, and six histidines, flanking the COOH-terminus of the coding sequence of LSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of LSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five column volumes of wash buffer. LSP was eluted stepwise with various concentration imidazole buffers.

Example 4: Protein Fusions

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5'and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/1 of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

DEX0273_119 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
150-165	1.11	16

DEX0273_125 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
55-76	1.00	22

DEX0273_132 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
3-27	1.06	25

DEX0273_136 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
36-46	1.18	11

DEX0273_138 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
45-59	1.10	15
15-26	1.03	12

DEX0273_139 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
24-36	1.06	13

DEX0273_143 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
96-110	1.11	15
74-89	1.05	16

DEX0273_145 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
439-450	1.05	12
407-421	1.04	15
643-662	1.03	20

DEX0273_150 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
24-54	1.16	31
147-160	1.16	14

DEX0273_151 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
140-150	1.14	11
121-135	1.07	15

DEX0273_159 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
40-58	1.15	19
100-141	1.15	42
19-35	1.11	17
230-248	1.09	19
203-224	1.09	22
158-173	1.07	16

DEX0273_160 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
8-41	1.04	34

DEX0273_161 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
46-55	1.06	10

DEX0273_162 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
53-62	1.10	10
36-50	1.01	15

DEX0273_166 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
140-152	1.04	13

DEX0273_167 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
49-58	1.05	10

DEX0273_170 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length
25-41	1.09	17

DEX0273_171 Antigenicity Index(Jameson-Wolf)

positions	AI avg	length

39-94 1.11 56
DEX0273_173 Antigenicity Index(Jameson-Wolf)
positions AI avg length
23-57 1.22 35
DEX0273_175 Antigenicity Index(Jameson-Wolf)
positions AI avg length
101-120 1.15 20
80-99 1.09 20
DEX0273_178 Antigenicity Index(Jameson-Wolf)
positions AI avg length
64-73 1.12 10
DEX0273_179 Antigenicity Index(Jameson-Wolf)
positions AI avg length
38-48 1.07 11
DEX0273_180 Antigenicity Index(Jameson-Wolf)
positions AI avg length
25-36 1.21 12
DEX0273_182 Antigenicity Index(Jameson-Wolf)
positions AI avg length
21-50 1.02 30
DEX0273_186 Antigenicity Index(Jameson-Wolf)
positions AI avg length
9-34 1.13 26
114-123 1.02 10
DEX0273_187 Antigenicity Index(Jameson-Wolf)
positions AI avg length
77-93 1.17 17
DEX0273_188 Antigenicity Index(Jameson-Wolf)
positions AI avg length
156-181 1.07 26
11-29 1.06 19
54-94 1.03 41
DEX0273_189 Antigenicity Index(Jameson-Wolf)
positions AI avg length
45-79 1.17 35
87-99 1.16 13
DEX0273_190 Antigenicity Index(Jameson-Wolf)
positions AI avg length
4-27 1.20 24
DEX0273_194 Antigenicity Index(Jameson-Wolf)
positions AI avg length
6-119 1.10 114
DEX0273_196 Antigenicity Index(Jameson-Wolf)
positions AI avg length
138-157 1.06 20
87-99 1.05 13
206-237 1.00 32
DEX0273_197 Antigenicity Index(Jameson-Wolf)
positions AI avg length
4-21 1.11 18
55-67 1.11 13
DEX0273_198 Antigenicity Index(Jameson-Wolf)
positions AI avg length
36-47 1.10 12
DEX0273_199 Antigenicity Index(Jameson-Wolf)
positions AI avg length
43-52 1.18 10
DEX0273_201 Antigenicity Index(Jameson-Wolf)

	positions	AI avg	length
	127-139	1.14	13
DEX0273_202	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	168-183	1.08	16
	58-78	1.04	21
DEX0273_203	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	50-76	1.12	27
DEX0273_208	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length
	142-163	1.11	22
	79-123	1.03	45
	65-77	1.01	13

The predicted helical regions are as follows:

DEX0273_122	PredHel=3	Topology=o4-22i29-51o61-78i
DEX0273_125	PredHel=1	Topology=o10-32i
DEX0273_129	PredHel=1	Topology=i7-25o
DEX0273_130	PredHel=1	Topology=i5-27o
DEX0273_137	PredHel=1	Topology=i7-28o
DEX0273_146	PredHel=3	Topology=i30-48o52-71i97-119o
DEX0273_147	PredHel=1	Topology=i13-35o
DEX0273_149	PredHel=1	Topology=i7-26o
DEX0273_162	PredHel=1	Topology=i63-85o
DEX0273_169	PredHel=2	Topology=o4-26i178-200o
DEX0273_176	PredHel=8	Topology=i2-24o34-56i61-83o93-115i128-150o155-177i184-206o210-232
DEX0273_177	PredHel=3	Topology=i21-43o58-80i92-114o
DEX0273_182	PredHel=1	Topology=i61-83o
DEX0273_185	PredHel=2	Topology=o15-37i185-207o
DEX0273_192	PredHel=5	Topology=i13-35o50-72i79-98o108-130i137-159o
DEX0273_193	PredHel=4	Topology=i5-27o61-83i96-118o128-150i
DEX0273_195	PredHel=2	Topology=i7-29o39-61i
DEX0273_207	PredHel=1	Topology=i5-27o

Examples of post-translational modifications (PTMs) of the LSP of this invention are listed below. In addition, antibodies that specifically bind such post-translational modificatic may be useful as a diagnostic or as therapeutic. Using the ProSite database (Bairoch et al., Nucleic Acids Res. 25(1):217-221 (1997), the contents of which are incorporated by referenc the following PTMs were predicted for the LSPs of the invention (http://npsa-pbil.ibcp.fr/cgi

DEX0273_127	Ck2_Phospho_Site 7-10; Pkc_Phospho_Site 34-36;
DEX0273_128	Ck2_Phospho_Site 25-28; Pkc_Phospho_Site 25-27;
DEX0273_131	Camp_Phospho_Site 78-81; Ck2_Phospho_Site 21-24; Myristyl 47-52; Pkc_Phospho_Site 80-82;81-83;
DEX0273_132	Amidation 17-20; Asn_Glycosylation 72-75;90-93;101-104; Ck2_Phospho_Site 3-6;27-30;79-82; Pkc_Phospho_Site 3-5;73-75;78-80;79-81;
DEX0273_133	Ck2_Phospho_Site 9-12;
DEX0273_134	Pkc_Phospho_Site 24-26;
DEX0273_135	Ck2_Phospho_Site 4-7;
DEX0273_137	Myristyl 51-56;63-68; Pkc_Phospho_Site 96-98;
DEX0273_138	Asn_Glycosylation 54-57;
DEX0273_139	Ck2_Phospho_Site 16-19;23-26;31-34; Myristyl 32-37;
DEX0273_140	Asn_Glycosylation 9-12; Myristyl 6-11;13-18;25-30; Pkc_Phospho_Site 17-19;31-33;
DEX0273_143	Asn_Glycosylation 118-121; Ck2_Phospho_Site 19-22;185-188; Myristyl 108-113; Pkc_Phospho_Site 180-182; Tyr_Phospho_Site 182-189;
DEX0273_145	Asn_Glycosylation 287-290;344-347; Camp_Phospho_Site 252-255;710-713; Ck2_Phospho_Site 6-9;12-15;17-20;61-64;101-104;118-121;187-190;251-254;290-293;338-341;398-401;459-462;514-517;522-525;546-549; Myristyl 55-60;73-78;76-81;107-112;550-555;596-601; Pkc_Phospho_Site 94-96;210-212;251-253;289-291;406-408;567-569;568-570;571-573; Tyr_Phospho_Site 321-328;646-654;
DEX0273_146	Myristyl 37-42;39-44;136-141; Pkc_Phospho_Site 27-29;67-69;76-78;161-163;
DEX0273_147	Leucine_Zipper 6-27; Myristyl 14-19;
DEX0273_148	Amidation 20-23; Ck2_Phospho_Site 16-19;
DEX0273_149	Myristyl 21-26;
DEX0273_150	Asn_Glycosylation 47-50;157-160; Camp_Phospho_Site 60-63; Ck2_Phospho_Site 27-30; Myristyl 155-160; Pkc_Phospho_Site 46-48; Tyr_Phospho_Site 130-137;
DEX0273_151	Camp_Phospho_Site 146-149; Ck2_Phospho_Site 109-112;155-158; Pkc_Phospho_Site 101-103;123-125;155-157;162-164;186-188;
DEX0273_155	Ck2_Phospho_Site 8-11; Glycosaminoglycan 42-45; Myristyl 44-49; Pkc_Phospho_Site 20-22;21-23;
DEX0273_156	Asn_Glycosylation 76-79; Ck2_Phospho_Site 21-24; Myristyl 35-40; Pkc_Phospho_Site 8-10;
DEX0273_157	Myristyl 49-54; Pkc_Phospho_Site 34-36;62-64;
DEX0273_158	Ck2_Phospho_Site 79-82; Leucine_Zipper 15-36; Myristyl 19-24;31-36;44-49;94-99; Pkc_Phospho_Site 12-14;26-28;89-91;
DEX0273_159	Asn_Glycosylation 148-151; Pkc_Phospho_Site 27-29;127-129; Prokar_Lipoprotein 18-28;
DEX0273_160	Pkc_Phospho_Site 44-46;
DEX0273_161	Myristyl 50-55; Pkc_Phospho_Site 32-34;47-49;54-56;
DEX0273_163	Myristyl 32-37;
DEX0273_164	Camp_Phospho_Site 27-30; Ck2_Phospho_Site 17-20; Pkc_Phospho_Site 11-13;14-16;30-32;
DEX0273_165	Asn_Glycosylation 45-48;50-53; Ig_Mhc 25-31;
DEX0273_166	Asn_Glycosylation 79-82; Camp_Phospho_Site 49-52; Ck2_Phospho_Site 24-27;37-40; Myristyl 66-71;88-93; Pkc_Phospho_Site 32-34;48-50;148-150;
DEX0273_167	Asn_Glycosylation 98-101; Camp_Phospho_Site 36-39;53-56; Ck2_Phospho_Site 85-88; Myristyl 58-63;66-71;72-77;109-114; Pkc_Phospho_Site 8-10;25-27;45-47;100-102; Prokar_Lipoprotein 63-73;
DEX0273_168	Asn_Glycosylation 45-48;50-53; Ig_Mhc 25-31;
DEX0273_169	Asn_Glycosylation 171-174; Ck2_Phospho_Site 78-81;90-93; Myristyl 57-62;60-65; Pkc_Phospho_Site 106-108; Tyr_Phospho_Site 119-127;
DEX0273_170	Ck2_Phospho_Site 9-12; Myristyl 44-49; Pkc_Phospho_Site 16-18;32-34; Tyr_Phospho_Site 30-36;
DEX0273_171	Ck2_Phospho_Site 56-59; Pkc_Phospho_Site 6-8;115-117;
DEX0273_172	Myristyl 9-14;36-41;67-72; Pkc_Phospho_Site 32-34;75-77;

DEX0273_173	Camp_Phospho_Site 26-29;27-30; Ck2_Phospho_Site 38-41; Myristyl 21-26;45-50; Pkc_Phospho_Site 24-26;25-27;30-32;34-36;38-40;
DEX0273_174	Ck2_Phospho_Site 15-18;67-70;104-107; Myristyl 57-62;76-81;87-92; Pkc_Phospho_Site 7-9;15-17;33-35;
DEX0273_175	Camp_Phospho_Site 96-99; Ck2_Phospho_Site 80-83; Pkc_Phospho_Site 47-49;92-94;102-104;106-108;
DEX0273_176	Pkc_Phospho_Site 232-234; Prokar_Lipoprotein 20-30;135-145;141-151;
DEX0273_177	Myristyl 83-88; Prokar_Lipoprotein 53-63;
DEX0273_178	Ck2_Phospho_Site 65-68; Myristyl 42-47; Pkc_Phospho_Site 28-30;
DEX0273_179	Rgd 11-13;
DEX0273_180	Myristyl 12-17;35-40;62-67; Pkc_Phospho_Site 75-77;
DEX0273_181	Ck2_Phospho_Site 25-28; Pkc_Phospho_Site 4-6;25-27;63-65;71-73;
DEX0273_182	Myristyl 11-16;16-21; Pkc_Phospho_Site 27-29;32-34;55-57; Tyr_Phospho_Site 6-14;7-14;
DEX0273_183	Asn_Glycosylation 20-23;47-50; Ck2_Phospho_Site 42-45; Myristyl 60-65; Pkc_Phospho_Site 8-10;48-50;89-91;90-92; Rgd 15-17;
DEX0273_184	Asn_Glycosylation 45-48;50-53; Ig_Mhc 25-31;
DEX0273_185	Asn_Glycosylation 178-181; Ck2_Phospho_Site 85-88;97-100; Myristyl 64-69;67-72; Pkc_Phospho_Site 39-41;113-115; Tyr_Phospho_Site 126-134;
DEX0273_186	Asn_Glycosylation 15-18; Ck2_Phospho_Site 18-21;61-64;129-132; Myristyl 33-38;74-79;119-124;120-125; Pkc_Phospho_Site 52-54;61-63;
DEX0273_187	Camp_Phospho_Site 56-59; Ck2_Phospho_Site 46-49; Myristyl 23-28;72-77;83-88;84-89; Pkc_Phospho_Site 59-61;78-80;88-90;
DEX0273_188	Amidation 20-23;160-163; Ck2_Phospho_Site 13-16;103-106;166-169; Myristyl 24-29;97-102;127-132;137-142;157-162;197-202; Pkc_Phospho_Site 39-41;73-75;103-105;110-112;132-134;166-168; Rgd 163-165;
DEX0273_189	Amidation 64-67; Ck2_Phospho_Site 72-75; Glycosaminoglycan 54-57; Myristyl 84-89; Pkc_Phospho_Site 16-18;46-48;72-74;88-90;
DEX0273_190	Pkc_Phospho_Site 5-7;
DEX0273_191	Camp_Phospho_Site 10-13;107-110;108-111; Ck2_Phospho_Site 78-81;100-103;111-114;132-135; Pkc_Phospho_Site 8-10;13-15;63-65;111-113;142-144;
DEX0273_193	Ck2_Phospho_Site 48-51;87-90; Leucine_Zipper 109-130;116-137; Myristyl 94-99;129-134;
DEX0273_194	Asn_Glycosylation 67-70;81-84; Camp_Phospho_Site 43-46;51-54; Ck2_Phospho_Site 2-5;29-32;46-49;104-107; Pkc_Phospho_Site 29-31;40-42;46-48;54-56;55-57;66-68;104-106; Tyr_Phospho_Site 8-16;
DEX0273_195	Ck2_Phospho_Site 83-86;87-90; Pkc_Phospho_Site 67-69; Prokar_Lipoprotein 7-17; Tyr_Phospho_Site 62-70;
DEX0273_196	Camp_Phospho_Site 297-300; Ck2_Phospho_Site 137-140;139-142;180-183;227-230;268-271; Myristyl 9-14;91-96;302-307; Pkc_Phospho_Site 20-22;95-97;139-141;150-152;169-171;197-199;227-229;268-270;275-277;305-307;
DEX0273_197	Ck2_Phospho_Site 104-107; Pkc_Phospho_Site 21-23;31-33;41-43;56-58;80-82;
DEX0273_198	Camp_Phospho_Site 27-30; Ck2_Phospho_Site 36-39; Myristyl 57-62;
DEX0273_199	Asn_Glycosylation 77-80; Leucine_Zipper 81-102;
DEX0273_201	Ck2_Phospho_Site 129-132;141-144;278-281; Myristyl 57-62;66-71;74-79;212-217;244-249; Pkc_Phospho_Site 120-122;128-130;129-131;203-205;224-226;227-229;256-258;338-340;
DEX0273_202	Camp_Phospho_Site 60-63; Ck2_Phospho_Site 130-133;209-212; Ig_Mhc 200-206; Myristyl 19-24;28-33;71-76;75-80;109-114;116-121;167-172; Pkc_Phospho_Site 66-68;196-198;
DEX0273_203	Asn_Glycosylation 48-51; Myristyl 98-103;128-133;133-138; Pkc_Phospho_Site 2-4;69-71;110-112;
DEX0273_204	Pkc_Phospho_Site 10-12;43-45;
DEX0273_205	Amidation 110-113; Camp_Phospho_Site 5-8;6-9;44-47; Ck2_Phospho_Site 51-54;67-70; Pkc_Phospho_Site 8-10;9-11;47-49;94-96;

DEX0273_206	Asn_Glycosylation 8-11; Ck2_Phospho_Site 53-56; Myristyl 31-36;32-37; Pkc_Phospho_Site 20-22;41-43;53-55;
DEX0273_207	Myristyl 36-41; Pkc_Phospho_Site 21-23;44-46;
DEX0273_208	Amidation 57-60; Asn_Glycosylation 3-6; Camp_Phospho_Site 59-62; Ck2_Phospho_Site 20-23;128-131;153-156; Myristyl 122-127;124-129;125-130;

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 115. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al., Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al., Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al., Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Manheim), and FISH is performed as described in Johnson *et al., Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and

chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 μg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 μl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 μl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of

administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 , μg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μg/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad.

Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 μg/kg of the polypeptide for six consecutive days. Preferably, the

polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5'and 3'end sequences respectively as set forth in Example 1. Preferably, the 5'primer contains an EcoRI site and the 3'primer

includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-*In Vivo*

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known

in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by

the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 μg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about

0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989); etc. For a review of such techniques, see Gordon,"Transgenic Animals," Intl.

Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated [redacted] (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in

(rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such

approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the

development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 116 through 208;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 115;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or

(d) a nucleic acid molecule having at least 60% sequence identity to the nucleic acid molecule of (a) or (b).

2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.

3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.

4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.

5. The nucleic acid molecule according to claim 4, wherein the nucleic acid molecule is a human nucleic acid molecule.

6. A method for determining the presence of a lung specific nucleic acid (LSNA) in a sample, comprising the steps of:

(a) contacting the sample with the nucleic acid molecule according to claim 1 under conditions in which the nucleic acid molecule will selectively hybridize to a lung specific nucleic acid; and

(b) detecting hybridization of the nucleic acid molecule to a LSNA in the sample, wherein the detection of the hybridization indicates the presence of a LSNA in the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 116 through 208 ; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 115.

12. An antibody or fragment thereof that specifically binds to the polypeptide according to claim 11.

13. A method for determining the presence of a lung specific protein in a sample, comprising the steps of:

(a) contacting the sample with the antibody according to claim 12 under conditions in which the antibody will selectively bind to the lung specific protein; and

(b) detecting binding of the antibody to a lung specific protein in the sample, wherein the detection of binding indicates the presence of a lung specific protein in the sample.

14. A method for diagnosing and monitoring the presence and metastases of lung cancer in a patient, comprising the steps of:

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the lung specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of lung cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.

16. A method of treating a patient with lung cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein said administration induces an immune response against the lung cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

SEQUENCE LISTING

<110> Macina, Roberto
Recipon, Herve
Chen, Sei-Yu
Sun, Yongming
Liu, Chenghua
diaDexus, Inc.

<120> Compositions and Methods Relating to Lung Specific Genes and Proteins

<130> DEX-0274

<150> 60/252,054
<151> 2000-11-20

<160> 208

<170> PatentIn version 3.1

<210> 1
<211> 402
<212> DNA
<213> Homo sapien

<400> 1
cgtggtcgcg gcgaggtaca actctgagat ggtaacttac tctccagagc tcccctctgg 60

gattaggctg aggttatcct gtgtgggaaa ggcttgaggt cacactctta tctggcttcc 120

tcgccttcac tttcctgctt acttcaccct tttattagga tcactaataa gcatttactt 180

gataaatcac tggcctatga accttcatct cagggtctac ttctgggaaa actgacctag 240

aagacaaatt atacaccaag gcctaatggg tgagctcatt attccctctt aaataaacac 300

ttagtttttt tcaagtattc aacatggcat gtgcaccttc ctactttgtt ttgggtgaaa 360

aaaaaatcag caaccaacat tacatgccct ctgttattat at 402

<210> 2
<211> 236
<212> DNA
<213> Homo sapien

<400> 2
cgtggtccgc ggcgaggtgc cgtttgaggc tagtttttta aggcacaact cagaccctga 60

ttagactgga taggaacaga tcttgaaaga atcctattaa tgatacttga tatctgtcta 120

tacgctataa tggcctacgt tatgatcatg aattagtcca tgctaaaatg gccagactaa 180

ggtggtaacg gggaattaaa caagctggac atggataggc cgtggatgtc gccaca 236

<210> 3
<211> 210

<212> DNA
<213> Homo sapien

<400> 3
ggtcgcggcg aggtgaaagg actgcttgag cccgggaggc tgaggctacg atgacccatg 60

tttgtgccac tgcactccag cctgggcgcc aaagcgagac cccgtctcaa aaaacaaaaa 120

caaaacaaaa tgaaacaatt aacaaagtaa cagacaacct acagaatggt agaaaatatt 180

tgccaactat gcatccaaca aagatctaat 210

<210> 4
<211> 3193
<212> DNA
<213> Homo sapien

<400> 4
ctagtgctgg cagacactgg ctttttattt ttaggataag aaaacaggca tattctttgt 60

ggtccattat ctagagccca tacttgggca gcatttgaaa tttcaccttа accacagaca 120

gggctccagg gaagtggaga tgtaattctt acaacaacag ttctgatcat ggccatggtg 180

atgactttcc aggtctcgtg ttcaagtggt gccagaatgc aggagccggt gggcagccct 240

gaggggttgc cttggccgca gcctctgtgc acgctcttcc tggtgtcctc ttacccggta 300

gctgtgcgct tgttcccgtg agaacagcct gcttccagag tgcccaggag tgctggtcag 360

ggacagtgcc cgtgaggctg cagaggaggt ggggtccatg gcccacccat ctctccctcg 420

ccagcagccc tggccagtgt catcctggtg tagaaagggt tgcgcacagg ataggaggga 480

gccacagttc ttgcttagct gtgctcacga ccggcttgca gtcctgtgtt tcttaagatt 540

gtatttggaa tggtaatatc cttagaattt tgggatattg agcttcatgg attttctctc 600

caaaacaagc cagcacaact aactgtagca gaattgtatc cactcattca ttcaactgag 660

atgaagtgcc ctccctttc cagggcctgg gctagttcct ggaatgcaac agagatttcc 720

gtggacacag tctctagtct cacagagcat atagtctagt ccacgattgg caagctgcaa 780

ccacagaccc agtccggcac tctgcccatt tttctaagtg aaattttctt ggaatacagt 840

cacagctgtt tttttaacat ggtgtcttgg ctactttcag gctgcgacgg gagaattgaa 900

tagttgtgac agagaccaca ggcccactga aagggacaaa cagggctgaa aatactcact 960

gtttggccct ttccagaata gcaagtttgc tggcccttga gctagcctgc ctttatgggg 1020

tttttttgt ttgttttttt aagctttcag cttcatgctg ctgtattttt agttgaagtg 1080

ttctgagtaa cagtcagtgt ataaaagggg attgcagaaa aaaatgaggg cttgctttac 1140

tcaacagaaa atatggccct tcctgaatga cactaggaga gtcattttat ctcatacatt 1200

cccttcattt cgttggtgga catttgttga aaccggcact caatggtcaa accgtctgtg 1260

ccctccagtt gctgacagtc ctgcaggaag atggacaaga ggcccagtgc tgacagtcac 1320

acgactctca ctacttgaat gaggggactg tgggtgcaac tagaaaatat gttgattctt 1380

agccattccc accttgcctc tccgttcaga accccagctg cgagctgttt gtttccctgc 1440

ctggaaatga tgttttaggc aggttcctta atttctcagg tctgtctcag ataataaaaa 1500

gctctttgta tgagcctcag aactgtctct tcagtgaatg aaattaccag tcattatacg 1560

aagggacttt aaaaaatttg tggaaatact gaagtaaaag atgataaaaa aataaaaact 1620

tcatttcttg gctgggcaca gtggcttatg tttgtaatcc cagtactttg ggaggctgag 1680

[redacted]

tgtagtccca gctacttggg aggcagaggg aggaggatta cttgagccca ggagtttgag 1800

gttgcagtga gctgtgatca caccactgca ctctagcctg gacaacagag caagatcctc 1860

tctcttaaaa ccaccaacaa tgacaacaac aaaacaacat ttttatttct caatgtaagc 1920

tccatcaagg tcaagatact tttgtaagct gtgacaccag ccatttagtc cacctctaaa 1980

gaattgcggg ctctgggaat ttaaccatgt cagtgcagcc tttttaacat tattaacgga 2040

agaaaaaatg agtgctttta aagatttttt aaaatgagga aacaaagtca gaaggagcaa 2100

aatcgggact gtaaggtgga tgcctaatga tttcccaaca aaactcttga agaattgccc 2160

ttatttgatg agaagaatga gccaggagca ttgtcatcgt ggagaaagac actggtgagg 2220

ctttcctggg tgtgtttttg ctaaagcttc ggctaacttt ctcaaaacac tctcataata 2280

agatgttatt gtggccagat gcggtggctc acgcctgtaa tcccagcact ttgggaggct 2340

gaggtgggca gatcacgagg tcaggaaatc gagaccatcc tggctaacat ggtgaaaccc 2400

cgtcccgtct ccactaaaaa tacaaaaaat tagctgggcg tggtggcagg cacctctagt 2460

cccagctact ccagctactc ggaaggctga ggcaggagaa tggcgtgaac ctgggaggca 2520

gagcttgcag ctagctgaga ttgtgccact gcactccagc ctgggcgaca gagcgagact 2580

ccatctcaaa aaacaaaaac aaaacaaaat gaaacaatta acaaagtaac agacaaccta 2640

cagaatggta gaaaatattt gccaactatg catccaacaa agatctaata tccagaatct 2700

ataagaagct tcaaaaaatt tacaagcgaa aaacaagcaa cccattaaaa agaaagtggg 2760

caaagaacat gaacacattt caaaagaaga catatatgca tttaaaaagc atataaaaat 2820

cactcatcat cactaatcac tacagaagtg cataccatct cacaccagtc agaatggctg 2880

ttactagaaa gtcaataaat aacagatgct ggcaaggttg tggagaaaat ggaacacata 2940

cactgttggt gggagtgtaa attagtttag ccactgtgga aagcaggttg gtgattcctc 3000

aaagaactca gaattaccat tcaactcagc aatcccatta ttcccaaagg aatgcatatc 3060

ccaaggaaat ataaatcatt gtaccataaa ggcacatgca cgtgtatgtc cattgcagca 3120

ctgttcacaa tagcaaagat aaggaatcaa cctaaatgtg cattaataat aggctggctc 3180

gtgccgaatt ctt 3193

<210> 5
<211> 814
<212> DNA
<213> Homo sapien

<400> 5
gcgtggtcgc ggcgaggttt ttttttttt ttttttttt ttttttttg gatcaataaa 60

accacccca cttgttgttt tttgtgggaa accccaattt tggtcctggg gttaacccct 120

ttgggagtct cccaaggtgt tggtcttccc cggggtaacc ccaaagatat gggtcccatt 180

cccttattta aacaatttta aatctgtgtt ttaggggac cagcctatca acatcgtgtg 240

tttcttacac tattgggggg atttatgttt ccacccctat aaagatgggt tttatgctct 300

atgtgatagc ctccttggaa aatataatgc tggcccctat ataaacaata acacacaaca 360

aataccgcgc taatagagtg ggccccaaat tacaggagaa gccccacgat ggtcgatcaa 420

caccaatcta acacctcgtg gacatatgtc acacatctgt atctacacaa aaaaactagg 480

gcggcgcaca tactactcac cccccacctc tggtgtgcgc caacgaggag agcgagaagg 540

gacaccacac cagagagtgc ccacgccagg agaagacacc gagcggatac ccacgccaga 600

agatcgacaa ccacgcaggc acatatacgt ggggcacaac aaagacacac aagagaatgc 660

ccatcatagt agcaactacg caagaaggag aagaagaaag aaagaaccca gcgggcgcac 720

aggcgagacc aacctgcgac actaacaggg cgcgacacta cccctgcagt ggaccaacta 780

gatccaccac ggacgaagaa acaagaaccc tggt 814

<210> 6
<211> 189
<212> DNA
<213> Homo sapien

<400> 6
caagtgcatg taaacttgtc aaagtaagtg tgtgagggct cactgcttat cacccctagg 60

ttatcagcag tgagccctca cacacttact ttgacaattc atcatcttgt tctatattcc 120

ccttcaagag gtccatccag ttttggccca tccggggaat ctaagggaga ttattcatct 180

aggaatcca 189

<210> 7
<211> 475
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (428)..(428)
<223> a, c, g or t

<400> 7
gccgcccggg caggtcccag ttcatatgtg acatcttttt aaaaaaaata acaacaaaaa 60

aaaaatgaga gaaaagctaa aaaaaaaaaa gtaagggttg accggttatg ggtttccatc 120

ccacatacaa tatctgttta aaaggattcc ctgtaaaatt agtttaaagg gttttggccc 180

tagaaatccc gtagttctac tccttagagc actcacgcca tgggtctttc ccttccccgg 240

ggttttaaac cttcatatac cttccagaaa tttgggagag caaaattttt ggcttggtcc 300

actggcacta tcatttataa aaaagctggg cgtaattcca tgggcctatt agctgttccc 360

ctggtgttga atattggttt atcccggctc cacaaattcc ccacacacaa acttaaccgg 420

gaccaaangg aaagacaaca aaaaagcgac acaacacgac gaaaacaccc agaca 475

<210> 8
<211> 622
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (412)..(499)
<223> a, c, g or t

<400> 8
cctatgtgat ggatcgcgga cgaggtacca gataatcctt acatgatatc ctggaaagcc 60

cctcaggcct gagtcaaatt gggatggctg gtcccccagc atgacccaaa caagcatttg 120

ctagcttagc tttacaacac agatgatgct atgggccaca gcaacttgag gacttgcctg 180

agccttgttc caggttaatt agacgttgct aaaagggtgg gctcattgtt aagtttggtt 240

tctaactaca ttactaaaat tagaaacctt aatataactt tcttctatag ttcaataacc 300

tggatgaggt atatctgccc tgcttataag atgtacacat tatgtagcaa aatggattga 360

agcagatggg ttaagagtaa gggtcttgtg tgttatgtgc tacataggcc cnnnnnnnnn 420

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480

nnnnnnnnnn nnnnnnnnna gcattgtatt ttcacaacgt aggaacaaga aaaaaaacaa 540

aatacaaata gacatgacac acaaagacac aacacatcaa ttaaataaat agaaactaat 600

acgcacataa aaaaattgta aa 622

<210> 9
<211> 799
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (589)..(676)
<223> a, c, g or t

<400> 9
ttttttttt ttccccagac cacttccaat gctggattac gtctcctcca aatgtgtatc 60

tggagagtga taatagtata ttaatttcat gggaagtggt ctggggaaaa agtaacaaga 120

aatctaataa aaaacataac tcatagttgc tgatatgata aatgataaat ttgatatgag 180

agaaagcagc aggttatatt tgtaaccaat tatccttaca tgatatcctg gaaaacccct 240

caggcctgag tcaaattggg atggctggtc ccccagcatg acccaaacaa gcatttgcta 300

gcttagcttt acaacacaga tgatgctatg ggccacagca acttgaggac ttgcctgagc 360

cttgttccag gttaattaga cgttgctaaa agggtgggct cattgttaag tttggtttct 420

aactacatta ctaaaattag aaaccttaat ataactttct tctatagttc aataacctgg 480

atgaggtata tctgccctgc ttataagatg tacacattat gtagcaaaat ggattgaagc 540

agatgggtta agagtaaggg tcttgtgtgt tatgtgctac ataggcccnn nnnnnnnnnn 600

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 660

nnnnnnnnnn nnnnnnagca ttgtattttc acaacgtagg aacaagaaaa aaaacaaaat 720

acaaatagac atgacacaca aagacacaac acatcaatta aataaataga aactaatacg 780

cacataaaaa aattgtaaa 799

<210> 10
<211> 344
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (55)..(304)
<223> a, c, g or t

<400> 10

gcgtggtcgc ggcgaggtac ttacttcaag caaataaatg cggtggctcg tgccnnnnnn 60

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300

nnnntgaaaa taaaataaaa tatatattta caggcctaca actt 344

<210> 11

aaaacaatat accaaaaatc tagatttttt cttcacatac aacaaatctt acccagggtt 360

tttggttcaa atacctgtct catttctttc cacatccgaa gttcttataa tcggtaaaca 420

taatactagc tactg 435

<210> 13
<211> 469
<212> DNA
<213> Homo sapien

<400> 13
cattctaggg tttcctttga gagaccggtc actgctgtcg caagtctcag ggagatggta 60

taatccctca cgttatccca gagtttttat aaaaatattt ctgagattag atggctacca 120

agagcgttca aatactttcc cttaatttta tcccacagtt tgttacttgc tttctaccac 180

tacttgagat gctattaggg tgtgcacatt tcctataggt gactttcgca atccgggcaa 240

gatggggctt tactctgaaa gactatctac tgggggggagg tgtgagggaa cagaaattct 300

ttcaaaagct gcccaaagag gtgttcaaag tttttgtccc tatcttccag tgtgttagcc 360

cggttcaccg atgctggatt tggtgggggc ccaggcgggt ttatataata cccaaatccc 420

gggcgaaaca tcttcccttt ggaactttct caatcctctt tgcacggga 469

<210> 14
<211> 741
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (607)..(607)
<223> a, c, g or t

<400> 14
tcgagcggcg cccgggcagg tacatgggct atgggcctct tgaggctgtg tcatgagcca 60

tggtcattca tattcagctc agaataaacc tcttcaaata tttgataagg tctagaattt 120

ttcctcaaca ctgcagatgt gctatcttct tataaaaaaa tctgaattat accaattctg 180

tagaagtgta ttaatctttc ctgcatacag aaaagattct ggtgtctttt tctatattat 240

caacaaacaa catattaata tctatatgtt gcacacagcc attatttcaa tacagctaaa 300

gataatattt caaaaattat agagaaagaa caacagaaat gaagaaagtt tttctatcct 360

tttgttttat attcttagaa taaactagaa aactctgtta ttactcctta cacaggtaga 420

atatgttgtg tatatttctt ttaaggcaaa aacatagcac ttgtgttttt tcaaacattt 480

tctttggatt aaatatgttt ttatcaataa taaaaacctt ttatcacaga ggttttacaa 540

acaaaaaaaa aacaaaaaca acaaaacaac acaaaaaagg gtcgggggggg ggaacaccct 600

gtggcgncaa acagcgcgtg tgtctccccc tgggggtgcg acatttgtgt tctccccgcg 660

ccccacaatt ccccccccaaa tttgcgcaca cacaaacccg aaccacaccc cgcccccaca 720

cccgccccgc ccccccccc g 741

<210> 15
<211> 850
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (716)..(716)
<223> a, c, g or t

<400> 15
acacgagggc actacacttg aagatatctg gctgaattaa tgttcacctt ccatgtatta 60

atttatgtct tcgcttataa ttcctatctc cctaaaatgt ataaaattaa actataactt 120

gactacctcg ggcactttct caggacctct tgaggactgt acctgagcca tggttattca 180

tattggctca gaataaacct ctttaaatat ttgataaggt ctagaatttt tcctcaacac 240

tgcagatgtg ctatcttctt ataaaaaaat ctgaattata ccaattctgt agaagtgtat 300

taatctttcc tgcatacaga aaagattctg gtgtcttttt ctatattatc aacaaacaac 360

atattaatat ctatatgttg cacacagcca ttatttcaat acagctaaag ataatatttc 420

aaaaattata gagaaagaac aacagaaatg aagaaagttt ttctatcctt ttgttttata 480

ttcttagaat aaactagaaa actctgttat tactccttac acaggtagaa tatgttgtgt 540

atatttcttt taaggcaaaa acatagcact tgtgtttttt caaacatttt ctttggatta 600

aatatgtttt tatcaataat aaaaaccttt tatcacagag gttttacaaa caaaaaaaaa 660

acaaaaacaa caaaacaaca caaaaaaggg tcgggggggg gaacaccctg tggcgncaaa 720

cagcgcgtgt gtctcccccc gggggtgcga catttgtgtt ctccccgcgc cccacaattc 780

ccccccaaat ttgcgcacac acaaacccga accacacccc gcccccacac ccgccccgcc 840

ccccccccg 850

<210> 16
<211> 616
<212> DNA
<213> Homo sapien

<400> 16
aggcagtgtc tgcgaagtca ataacacaca taggtgtgcc atcaggtgtc ccgttttgcg 60

ccagtagaag cctcgacgct ctctcagtgt ctctggctat ttaggctgac aaggcaaact 120

agtagaagtc tctctttacc caagacccag tgtagaagtg aaactctggc atttgagatg 180

tatacttttc tggcctcatt ttgagacttt tgaatatccc catcacgata ttgattattt 240

tttgccggca gtcctaagaa gggatgattc acgggtctgg gggaaaaccg ccagccacga 300

gttcatgggc agtaagattg gtggaccgac gctgtggttc aagaattccg aaatctattg 360

acctgcaggt ttggaagggc ctcttgcaag agcctgggcg tagtcctagg ccataggctg 420

gtccccgggt gtaactgtgt atccggccca gtccaaaagg ccaagagcaa cccggacccc 480

acagactcga ggcagcagcg cgtagagaat accgataaca accaagacga gaggctacaa 540

cacgagagca gaaacgagga gataacacaa aacgagagac ccacagagga cgaaaaagta 600

acagagaaac acagcg 616

<210> 17
<211> 876
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (206)..(206)
<223> a, c, g or t

<400> 17
ccaccccagg gggggggggg cggggacctt aaacttacag gtcccaggaa aatgggtgtt 60

gggatcccag ttctttggca cattggtctc tctcctgaaa ataaacatct ccaaacatat 120

taacactcct ctctctaaag cctggggtgt aattccatgg gtccataagc tggttccctg 180

gtggtggaac atgggtgctc tccccncgcg tccacaatat ctccccacca caaacatata 240

caccgagaca caaaagagag gacgaacgaa gaggaaaaca gcgcaggaca ccgcccgaaa 300

acaagggagc cccgccaagc accacaaaga aaacaaaaga aaccgaacag gacaaggagc 360

gacaacccag aacagagaac aagagggaag aagagaaaca caaacaaaca gacggaaaac 420

aaacggaacg caagagacac acggaacgac gaacaggcga agaagagacg ccaaccacac 480

aagacagaaa agcgaacacg aaacaaaacc gccgcaggaa cccacagcga cccagaaaac 540

gcaacacaag accagcaaaa cagcgcagag gcacgcaacg cagaagacgg caaggggcac 600

caaggagaaa acaggagacg acggaggcgg cgggcaagaa acgacacaag agggaagagg 660

acgaggggaa gggaagaggg aaagacaagg cgagcagaga aagtcagcag aaacaacgga 720

agaaagacac cgaaaccacg acagcgaaaa gaacacaaga caagcagaca agacaaagaa 780

gaacggaaca agcaaagaac gaacacaaaa gcacagaagc cagagcaaca gagaaagaga 840

agaaacagaa acaaagaaga aggacgagag gcgaac 876

<210> 18
<211> 474
<212> DNA
<213> Homo sapien

<400> 18
cgtggtcgcg gcgaggtacc gaaggtgtca gtgttgggga tggagagtca tagtggtgat 60

aagcctggta agtaacctca ccaggacgtg ccaaagacag gtcagcgagg tgaagggact 120

gtggaagcaa ggtaagggga ggtgaagttg tagtggaatt tgggaggtgc ttactgatct 180

tcttgcaggc cctacaaatg ttattcaaac ctctgggcaa atgtattagt cacttgaccc 240

tccacgaact cctccaagga cttcagggat taacgctgtt gccacctggc tcctcagagc 300

ggccagtcac cgtggtgctt cagaatcagg taacatgtct aggaggcttt tttccataga 360

tggctttcag gttggtatca ctgataaggg gtaagttggg ggacagtctc atctctacac 420

aaatcttatc ctctgcagtg cttctctatt tctagtaagc acatgatcac ctgg 474

<210> 19
<211> 563
<212> DNA
<213> Homo sapien

<400> 19
ggtgtcagtg tgctgaggag gtctgagtaa aagattagga cgccctgata aatgttgagc 60

ccctatctac ttttcttaga gaaggattcc taggacaagt aggtaccgaa ggtgtcagtg 120

ttggggatgg agagtcatag tggtgataag cctggtaagt aacctcacca ggacgtgcca 180

aagacaggtc agcgagggaa gggactgtgg aagcaaggta aggggaggtg aagttgtagt 240

ggaatttggg aggtgcttac tgatcttctt gcaggcccta caaatgttat tcaaacctct 300

gggcaaatgt attagtcact tgaccctcca cgaactcctc caaggacttc agggattaac 360

gctgttgcca cctggctcct cagagcggcc cagtcaccgt ggtgcttcag aatcaggtga 420

acatagtcta ggaggctttt ttccatagat ggctttcagg ttggtatcac tgataagggg 480

taagttgggg gacagtctca tctctacaca aatcttatcc tctgcagtgc ttctctattt 540

ctagtaagca catgatcacc tgg 563

<210> 20

<211> 630
<212> DNA
<213> Homo sapien

<400> 20
aggatgatcg atcatatggg cgcatgggtc tctagatgct gctcgagcgg cgcagtgtga 60

tggattggtc gcggccgagg tacttccttt atccagacat aaatttaatg tgttgcaatc 120

tatttgacat gatttcttac aaaatttaag tttgtgggtt aagtcttatt tttagagatc 180

aatgctgata cttataaaat gccacttgaa aagatttcag ttgtgttgct taataccaaa 240

tattgcctac tttttgcaac atatttaaaa ataaagtaga aattcagctt cttaatacaa 300

atgtatgttg tttaatgaag caaaagtgaa gagactgaat tgttaattta ttttctagag 360

tgtctccaca ttcaaatggg cggatgatca ttggaaggtg gagggcatat taaataaaag 420

gcatttccat ctgcctatag ttgccagtta tctcaggaag ttagtgcatt gttttaatga 480

ggttacagtt tctggctaga tttccctagt gaggttagtg ctatttgtgc cacagagtgc 540

atttgccagt cattttacca ctgtgtctca attttgagta gagggcaaga ataaatcatt 600

taatttattc ttaaaacctg gggaaaataa 630

<210> 21
<211> 538
<212> DNA
<213> Homo sapien

<400> 21
tgctcgagcc gcgccatagt gatggatgcg gccgaggtac cctacatcaa agtctgcatt 60

caggtgatta taatattccc tcgtgcccat gccgaagaat gtatcacaga gaaattgtgc 120

ctgtttatga ggttctttcg gtgataactg gccttcaaat tcaggttttc agtggcaagg 180

aagctgacag tgttataaag cggtctattg gttggggtcc attctttaag cccaggtgtt 240

acaacccttg aaaaaaaaat gagtcaaagt gttgttcatg tgaggtatcc taagagtaga 300

cacagaggct actacagtat actacgattg acatttaggc ctgatgtctc cgtcaggttc 360

ctttagactt tctcagattt tccttttcct tgaggacttc aatagttatg ggtagtgctg 420

gctgactgta tcctttcatc tatctcacca gaagtataat acttttattt cgtttgagta 480

taaattcttg caccctaaat aagttgtcct tagtcatttg tattagctaa caaaatac 538

<210> 22
<211> 197
<212> DNA
<213> Homo sapien

<400> 22

aaggaccagg aacccgtaga acaggaccgc gatgcagggc agataaccaa tagggatccg 60

acaccctgga cgagccatca cagaagatcg aacggcccaa gtccgaagtg gcgaaccccg 120

gcacagggac ttacaagata ccagcggtcc ccccggaggg ccccgaggcc gcccagaccg 180

aacaggggaa cgggaac 197

<210> 23
<211> 1059
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (414)..(414)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (426)..(426)
<223> a, c, g or t

<400> 23
gtgaatacac tcactatagg gcctgttgcc tctagatgct gctcgagcgg cgcagtgtga 60

tggatcgtgg tcgcggcccg agataccatg tagtgctgtg tcttctccca aaaagatgtg 120

tatttagctt aggaaagaaa tgcaaagtgt ggttgataaa atggctcatg aaagtgcagt 180

gagactgacc ccatcctgta ttcagggata ggccatccct ctctgccagt gaagagagac 240

actatcttta tatccgtaat accacgtata gactctgggc ttccctgtag tccccctggg 300

gatagtgtcc tccacccccc attagtgtat tagtgtatta ctcgtggtcg tgcggtgaat 360

gtcgtctgct gagtgaggtg gatgtcttgg tctaggttac tatttttgga acantaactg 420

gctaanaccc cttcggaaca cacaaaaaca gggcaggatg tatattttaa ttttttaaaa 480

tttaccattt tatttcacgt tattgtacca agctcatgaa atgttttact atttggtcag 540

aaaagtgaca ttatggcaca ttgcattcct aagatttaat acatggtttc tcagggaggt 600

tgaaatacag tatcctgaat cttaaatatt atagaactct taaacaattt tggcttagct 660

ggagaaggct ggggtatatt taagaatgta tgtgttctgc atatactcct ttaagaaaca 720

gattttccag gctggctgtg gtggctcaca cccattaatc cccaacaact ttgtggggga 780

agcgcccgag ggcagcgagg gattgcttga ggcccaggag ttcgagacag cttaggcaac 840

agagcaagac tgatctctat taaaaataat aaaaagaacc cgctttgaga taatagtgat 900

aacccctgac tcgtcatatc acctagacaa ttgagattcg acactggctg ggatacgaga 960

ccagttgccg acctgtttct ggttcctttc ggtggggacg tttaaggggc caggcttttc 1020

ccgtctctac ccgtggtaat cggtctggtc tgcgtgtca 1059

<210> 24
<211> 1052
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (114)..(114)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (151)..(151)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (284)..(284)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (447)..(447)
<223> a, c, g or t

<400> 24
gcgtggtcgc ggccgaggta cgtgccgcgg aatatgcccc gcttgcaatc gacatcatcg 60

gtgccaaggg acctacgcat ccatcgcaga tgaacggtgg tccgacggct tgancaacgg 120

gtcatcagga caaggttgta agtgagacca ngttttatag atagcttatg catattctcg 180

cggaggccaa ttacgtatga ctcggggtga tgtcagaatg agttccatct ctccgagttg 240

tgccaagggc ctgatgtgcg ttccgctcgt cagataagaa cttngttaga ccttgcgacg 300

acgaaatcca cacgactagt cgagaactaa ttctaggtca taacataaca tacatgacaa 360

aaccaaaaaa aaacaaaaaa aacaaaccaa cacaaaagcg cgttggcgcg tgtaaacacc 420

agatgggctc tatacacgcg tgtgtanacc ccttgtgtgt gtcgacatat gtgtgtgtac 480

tccccgcgct ccccacaaat actcccccca cacaaaacat atcccccggc acacaaacgg 540

caaacaaagg aagagaagag aggggaaagc aagaaagaga agacagcaga aacaaagaga 600

aagacaaaaa ggaaaggaga gaaggaaagc aggaaaaaag caagaaagaa caaaggaccg 660

aagaacaaca cagaaacaaa aaaaaagcaa agacgggacg aggaaaaagc acaaaacgaa 720

agaaaaggaa aagagaagca gagaggagaa ggaaaaaaga gagaagaagg aacgaaccaa 780

aaagaaaaca gagaaagaga cagaacgaaa gaaagcgaca agacacaagc aaagagagcg 840

acaagaaaag acagaaaaaa agacaggaga caagaagaaa cagaaaaaga aagaagcaga 900

acaacaaaga gggaaaaaag aaaatagcaa aacgcaaaca gaaacaacaa acggaagaaa 960

gaccggacaa aacgagagag gagaagagaa aggcacaaag aaagaaaaag agaaaagcag 1020

agaaagaaga caaccaaaag aaagaaagaa cg 1052

<210> 25
<211> 1124
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (186)..(186)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (223)..(223)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (356)..(356)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (519)..(519)
<223> a, c, g or t

<400> 25
tagctgcttc ctttctctct cgcgcgcggt gtggtggcag caggcgcagc ccagcctcga 60

aatgcagaac gacgccggcg agttcgtgga cctgtacgtg ccgcggaaat gctccgctag 120

caatcgcatc atcggtgcca agggaccacg catccatcca gatgaacggt ggtccgacgg 180

cttgancaac gggtcatcag gacaaggttg taagtgagac cangttttat agatagctta 240

tgcatattct cgcggaggcc aattacgtat gactcggggt gatgtcagaa tgagttccat 300

ctctccgagt tgtgccaagg gcctgatgtg cgttccgctc gtcagataag aacttngtta 360

gaccttgcga cgacgaaatc cacacgacta gtcgagaact aattctaggt cataacataa 420

catacatgac aaaaccaaaa aaaaacaaaa aaaacaaacc aacacaaaag cgcgttggcg 480

cgtgtaaaca ccagatgggc tctatacacg cgtgtgtana ccccttgtgt gtgtcgacat 540

atgtgtgtgt actccccgcg ctccccacaa atactccccc cacacaaaac atatcccccg 600

gcacacaaac ggcaaacaaa ggaagagaag agaggggaaa gcaagaaaga gaagacagca 660

gaaacaaaga gaaagacaaa aaggaaagga gagaaggaaa gcaggaaaaa agcaagaaag 720

aacaaaggac cgaagaacaa cacagaaaca aaaaaaaagc aaagacggga cgaggaaaaa 780

gcacaaaacg aaagaaaagg aaaagagaag cagagaggag aaggaaaaaa gagagaagaa 840

ggaacgaacc aaaaagaaaa cagagaaaga gacagaacga aagaaagcga caagacacaa 900

gcaaagagag cgacaagaaa agacagaaaa aaagacagga gacaagaaga aacagaaaaa 960

gaaagaagca gaacaacaaa gagggaaaaa agaaaatagc aaaacgcaaa cagaaacaac 1020

aaacggaaga aagaccggac aaaacgagag aggagaagag aaaggcacaa agaaagaaaa 1080

agagaaaagc agagaaagaa gacaaccaaa agaaagaaag aacg 1124

<210> 26
<211> 659
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (239)..(239)
<223> a, c, g or t

<400> 26
tcgcggccga ggtttttttt ttttttttt ttttttgtgg gtgtttaaaa gtttaagtta 60

ggatatgggc ccatatacca aaaagcctca agggacaaca aagcctgtgc ccctctctcc 120

tataaggggg tgcccctctc aagagcccct atttgtgtgt gttaaacact ctcagagagg 180

aaaagctctc gaactctctc tgtggagccc ttctcctctct ccctcacgag tgtgtgggng 240

aaaactgtgc ccgaggattg agaggataaa ctccgtggct taaaatctct tggtgtattc 300

cccaaatatt aatgccccca acacaaatat tgtggaatat caccaccact tatttaaaat 360

atacacttac acatatctcc catatttaac gcggtctcaa tgagaatgtg gtattcacgt 420

ggcacatatt ctcaccatat tacacatctc gtggcacata ctccacaaga agcaagcgcc 480

tttgggcgag ggggatctct tatattctac aagcctgtgg gggatatatc tcgatgtggc 540

gcccatataa gcgctgtgtg ttccgcggtg gtgtgtgaaa atgtgtggta tatctcgcgg 600

ctctcaccaa attctccacc acacaaaatt cgccggacaa caaaaaaggg gggggggggg 659

<210> 27

<211> 1337
<212> DNA
<213> Homo sapien

<400> 27
tttttttttt tttttttttt ttttaaagtg ggtaaaaatc tttatttatc tattttataa 60

attcacttgt gcaagaacaa cacttctcct caaaaatact tttccccccc aaaagagctt 120

aaaaaaataa gaaaaagagc taattagggt aggcagaaag tgtctcttgg gagacacccc 180

tctctgtgtt ttctcagagg gagaagcctc tagtgccggg cgtgtgtgtg tctccaacca 240

ccgagaggtc ttgtgccacc agaggggggcg agagagtctc tctccctgtg agacctctgt 300

gacacttgtg cgccagagac acctctctct gtgtggtgtt gtggcgcctc tcgcggagag 360

agacagcaac gccccaagct ctctgcgtgg gcggtgtgag agactctccg tttctcctct 420

cgagtctcag tgtgcgccca acacaggtgt tgtgtatctc tccactatat atagacgcca 480

tctctctcta taacacactt ttctcactct ctataagaga gatatatatc tcctatagag 540

tatataataa agatctctat actacccata tatattgtgt gagggcgcgc actatgtgtg 600

tgggtatatc tcccacagtt gggtgtttaa ccacacaaag aaacacatat aatctctatc 660

tctctctgtg ccatatatat tatgtgtgtg tgtagacatc tttatataag aggagaacaa 720

cagcgcatgt agagagaatg tgacctctct ctatatgttc tcacacacac aacacgtgtg 780

gggtgtgaaa tctctctcta tatgtgtgtg tctctcccac gaagttgtgt ctccccggtg 840

gggatggtgg ggggctctcc accccggaga caatgatgcc ccaatttctc ctctccctat 900

tctcgcgatg gatgcgccga gaataataat ttacaccata tatctctctg ttttttacac 960

acccatgttg tgggtgccca taaaggggag cgcggcaccc aaacatgatt agtgggagag 1020

agaatgtgaa aaaaaatata aacgaggccc gagggggggcg cagaataaaa ctacgagggg 1080

ggtccacaat agaagctccg aagatgtacc ccgccggggt ggttgcggca ccactattcg 1140

tggttgttat atccccccggt ctccccaccc atatttcccc cccccataat caattagaca 1200

gaacacaaac aacacaaaac acaacaaagc agactacaag caaaaaagac gaaccaaacc 1260

agcgacatag aaacaccacc aaccacaaaa caacgcacca gcaaaaccac acaacaccac 1320

acccatacag aaacaaa 1337

<210> 28
<211> 164
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature

<222> (111)..(111)
<223> a, c, g or t

<400> 28
acattgctaa ataacttctt aggaagagat gtggggtggc aaacccttgc acgtctgaaa 60

atatccagat agattcggct agtgtgtgag cacactgttg aaatgtcatc ntctccctgt 120

gactcttaca cggacactct ctctctattg tctataaacg cttg 164

<210> 29
<211> 183
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature

[redacted]

ggcataatca tggacaatag ctgtcccctg tgtgaatttg tttccgccac aattccatct 660

cacacaacaa aatggt 676

<210> 31
<211> 2040
<212> DNA
<213> Homo sapien

<400> 31
accattttgt tgtgtgagat ggaattgtgg cggaaacaaa ttcacacagg ggacagctat 60

tgtccatgat tatgccaagc ttctcttgaa aatatgcaat aaagaacaaa aaggaaaaag 120

gcctacataa aataccaacc tttgttaaaa tgaaatgcta taataaagag tagacactat 180

gaagtttcca caatggtaat tcacgttctc agtaccaaca taggcaaggg cacacagacg 240

tagaaattat aaagtaattt ccacaaagag ccctctttga atactaaaga ctgtagtact 300

tgctaaatac taaatactag aagccctatg aaatttttta tattaaggga gcaataaaaa 360

ttttttttt tgtttttgag acagagtctc gctctgtctc tcatggcctg gagatgcaac 420

taggcacaat actcggttca catgcaacat caaatatcct gggctcaaag atgatccctt 480

ccacactcag cctcttgaga tagactggga gcacacagat agtgcgccac ccacacacca 540

gactagatgt gtgtgtgtgt ctgtgtgtct gtatgtgtgt gtgtatgtgt gtagagactg 600

ggtttcgcca tgttgcccgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtctgtgtgt 660

gtagggactg ggtttcacca tgttgcccag gtgtgtgcat gtgtgtgtgt gtgtgtgtgt 720

gtagggactg ggtttcacca tgttgcccag gtgtgtgttt gtgtgtgtag agactgggtt 780

tcgccatgtt gcccaggtgt gtgtgtgtgt aaagactggg tttccccatg ttgcacaagc 840

tggtctattc tcaaactact gagctcaggc aatctgccca ccacagtctc ccaaagtgct 900

tggattacag gcagaagcca cagtgcctgg ccagcataaa ctattctaaa tagctttttt 960

tatttaacta ataaatctag acagattaaa cattttagag gacctctaaa atactatgcc 1020

ctgtggaaaa caagacaaag cactaattcc atacagcttg ccttgggaca gattctccct 1080

tcagtctcat ctgtgtaata cttattattc tcaaagaaag tgaacacata gagcgacatt 1140

taaattccaa gatgtaacaa aaccttaatg ttaacattaa aaaattaaaa tctcagagtg 1200

tgccacacca taggtgctta attaaaaaaa aacatactaa acagtgaaaa tggatgaccc 1260

agtccttagc ctatgttatg gagttagcga agcaagctcc agtgccctgt ggcttagtca 1320

tacaataaat acttactgtc acacagtggc tgctcagtaa atatttatgc tttttaaact 1380

aaacagtgaa aatgggtgac cagtccttag cctttgctta tgaagtgagc agaagcaaac 1440

tccagtgccc agtggcttag tcatacaata aatatttact gagcagctac tttgtgccac 1500

acactatgct aggttcttgg caacaaggac actgtttggt cattaaggaa acatggaaaa 1560

gtgagggatg ccccctctcc aagcaagcct gacccccctcc gcatggcctc caacacacgg 1620

ctgcttccac tctgggctgg caggtggatc tgtttacaga tgttatctct ctcatgaatc 1680

agctgcagaa cctgatgaaa cagaacacat tataggtaat cacaatctca ccaaagaacc 1740

ttacagaaag caataccgct cttactatgt atcctccaag gtcaattttc acataattaa 1800

gaggctaatt aaaccagaca cacaaaatca cctattccct aacttttgtt caagccccat 1860

tctatttgtc tcagacactt cacctgatgg catctctgct ttcaaagagt agagagaaga 1920

aagtaagcag aggtcagatt aaagccatgg gagctgaata caggtagtgc tgacactagg 1980

gtcagcaggc aaagcaggaa aaaaatggca cttctttcag ctagcttaca aagcagtcac 2040

<210> 32
<211> 285
<212> DNA
<213> Homo sapien

<400> 32
atgccgaccg gcgctagtgt gatggatgcg gcgcccgggc aagtactaca gatgggcgcc 60

accacatcca gctaattttt gtatttatgt tggttggttg gttttttgtt ttcgttttag 120

tttgtggaga gacaggtttt tgctgtttcc caggctattc taaagttcta ggctctgcct 180

gcatcagcct cccggggagc tgggattaca ggcgtgagcc actgtgccca gcccttagaa 240

ataattttct ccacctccat tcctctgact cttggtttgt gcctc 285

<210> 33
<211> 618
<212> DNA
<213> Homo sapien

<400> 33
ttccgagcgg cgccagtagt gatggatgtc gcggacgagg tgattttggt gatagaatta 60

caaaaactgc tagtggattt ttttttttt ttttttttt tttggaaacg gtttttgcct 120

ctgtcccccg gctggttgcg gggttgtctc ggtcttgacc cccgcccccc gggtgcagtg 180

atttccctgc ctcatttccc attgctggga ctacgggcgt gcaccaccac gcccagctat 240

ttttggtatt ttatagcaga gacagggttt cccagtgtgg gccgggcgtg gttctcgaac 300

tttccgaccc tcaaattgac ctccgccctc cttgggccct cccaaagttg cgtgggacta 360

ccaggcgttg agccccggcc gtggcctcca atatttccgt tgtccataaa ttccaacagt 420

tggccctccc tttgagccat cgaggtgtgg gggcaaaaaa aacatctttc ggttaatatt 480

aaaatgggcg ttctatccca tcacagacag ggcaaaggag ggggcgacaa aaagctggga 540

gtatccttgg gccataaggc tgttccctgg tgtgaaattg gttttcccgt ccacaatccc 600

cacacataac cagaccac 618

<210> 34
<211> 365
<212> DNA
<213> Homo sapien

<400> 34
aaaaaaagaa gaagttctgc aatttggatt tctccccata agttagacag gggaagaaga 60

tgagaaatta gaaaattcat acggagggga acagggggag aagcagaggt tactggggaa 120

actccttagg ggcaaaacaa ggcaggtctt atagaagggc tgggtcggct gtaacttctt 180

caagggtaaa ccaccaacaa taaagtctgg gggtaattca tggtccatag cctgttccct 240

gttgtgaaca tggtttatcc gctccacaat tcccacacaa tatctcggga agacagtcca 300

acgaaacgag taaaaaccaa gacaaccatc aaaacgaaca gaaaaaacag cagacaacaa 360

agaga 365

<210> 35
<211> 276
<212> DNA
<213> Homo sapien

<400> 35
accaaattga taaacagcag gattcctgcc ctgtggaggg tatgtgttca tcaaaggagc 60

ccacagcttc agagtgagat aaggaaaaga acgggaaact gggggagaaa ataccagggg 120

gcataatgca gactaagggt gggaggggca agtggagtgg tcaggaaagg ccagtctgag 180

gaaatgacat ttcatccgag tctcagagac agaggcttgg aaaacatata ttccaggtat 240

aggagacaac atacgcaaag tccctggggc aggaaa 276

<210> 36
<211> 506
<212> DNA
<213> Homo sapien

<400> 36
accaaattga taaacagcag gattcctgcc ctgtggaggg tatgtgttca tcaaaggagc 60

ccacagcttc agagtgagat aaggaaaaga acgggaaact gggggagaaa ataccagggg 120

gcataatgca gactaagggt gggaggggca agtggagtgg tcaggaaagg ccagtctgag 180

gaaatgacat ttcatccgag tctcagagac agaggcttgg aaaacatata ttccaggtat 240

aggagacaac atacgcaaag tccctggggc aggaaagagt ttggtacatt tgaggaccaa 300

atagaaaact ggtatggcct tggtttatca tggctgacat acaaagtcca ttgcagatct 360

gaagtgatgg cctagggaga gagcaggacc tggaatgcca cagaccccag atcatcttcc 420

gtatataagg tgggcttcag agtttagctt cctctctaac ctcagagtca ccaggaggaa 480

tcaggaagca atttcaccac tctcct 506

<210> 37
<211> 249
<212> DNA
<213> Homo sapien

<400> 37
acaggaaggg gtcaaggtgg agagcaggct agagggaggc tggcgagatg ggccaggtca 60

ccatggcatg ctccacactg ctgggtgtag gaatgcatca cggggaggtg ctgacacttt 120

cagggtagac agggaacgtg gactgccaca caccgactca gggaaaagcc aacagtccca 180

tatgtaaatt ttaaagttag ctttagaaaa taagttaaca gttatcagag caaaagtaag 240

gataaagga 249

<210> 38
<211> 406
<212> DNA
<213> Homo sapien

<400> 38
agatgcatgt gctcgtcagt gtctccgtcc gctacggtgt tgcgtctatg cgtggtttgc 60

acgctgtgcg tctgggtcat gccttgcctg ctgttgtcct cttgtacatc tcagcgcacg 120

aattactcaa tcacgaccta tgactgacgt caatgacggt gaagcggaat cttcatgcac 180

acatccatat gagggtcacg atgaatgtcg gctacagcga tgcgaggtag tggcacaaat 240

ccagggcgcc agacacagca ttggctgacg tggtgagtga taggtatctt acggcagggg 300

agcatctgtg agtacagtca ccacaacgct atgagcgtaa ctcaatgtgt acactagact 360

agttatcctt gtgttgaaac ttgtatatcc agctcacata ttccat 406

<210> 39
<211> 253
<212> DNA
<213> Homo sapien

<400> 39
aagaattttc tttagagagc aaaacatcat tttgtggcaa ttcagaggaa cagtgaagat 60

ttctagcctc agatactggt gtggaagaag tagcagagct taatgctaga tcggctaaca 120

tatttagggc ctgggagtca tagttgacga tggagttttc aggaagatca ttgtgagccg 180

ctgtggtatt ttctggttga acactattta tgctaatccc atcttcttga ccacctcttg 240

aaatttctga ttg 253

<210> 40
<211> 1198
<212> DNA
<213> Homo sapien

<400> 40
agtgagaaaa gaaactgaac aaaaaaggat tctgaagaaa tgttgaaagc aaagaagaga 60

gtttttccat tgagtccagc gtcaaatctg agagtgcagc ctaagaggaa ggccagcatg 120

ccccacatgg tgcagagtaa aaaggtgaac ttgtgccgcc cctttcccaa aagaactgct 180

tccagagcag acaacagctc ggactctcca acaactctta agttagttaa aggacagttt 240

cctcagaaaa gaaaaagagg tgcggaagtg ctgactgcac agtttgtaca gaaaaccaaa 300

ttggatagga aaaaccaaga agctcctatt tctaaagatg ttccagtgcc aacaaatgct 360

aaaagggcaa ggaaacaaga gaaatctcca gtcaaaactg ttccaagggc taagccacct 420

gtgaagaaat ctccacaaaa acagagagta aatatagtaa aaggcaatga gaaccccaga 480

aacagaaagc agctacaacc tgtcaaagga gaactgcttc aaagcttcaa tcagaaattt 540

caagaggttg tcaagaagat gggattagca taaatagtgt tcaaccagaa aataccacag 600

cggctcacaa tgatcttcct gaaaactcca tcgtcaacta tgactcccag gccctaaata 660

tgttagccga tctagcatta agctctgcta cttcttccac accagtatct gaggctagaa 720

atcttcactg ttcctctgaa ttgccacaaa atgatgtttt gctctctaaa gaaaattctt 780

tgcgaggtac atctgaccat gaatatcata gaggagttaa aactcaaaaa ggtgaattac 840

tacctaaccc atcttctgat aggaagagta attctggatc agacttaact gttagccaag 900

atgaagaaag cttggttcct tgtagtcagg cccctgctaa agcccagtca gcacttactg 960

aggaaatgct agaatcttca gatgcaagcc aaagctcttc tgtttctgtg gaacattcat 1020

atgccctgct ccttacagaa cattcaaaga aacatctaca ggagagagag atactaagcc 1080

ctctgtttcc caggaatggg acaaaaagcc ctgaagcagc aaccccagtg gggaaagtca 1140

tgccattcgg catcagccgg ctttgtgctt tcagcaaagc tcctgacgac ccgtggtg 1198

<210> 41
<211> 151
<212> DNA
<213> Homo sapien

<400> 41
ccgcccgggc aggtacctaa acaggccaaa tgttgccttt ggggttcctg tttcaacagc 60

atggtgtgaa gcgccgcatc aaccttctct gcctattaaa ataaaatgtc ataaactcat 120

cctgcaaggt ggcaaattcc tcaagaatat g 151

<210> 42
<211> 3096
<212> DNA
<213> Homo sapien

<400> 42
ttcctcacga aactcccagg cgctgtatag gaaacataaa tccgttgtca ggcagcagta 60

gcacgctgtt gctctcggag cttggctgct cgttcgtgct cgcaaccact aaggtctacg 120

caaacctcca cggtttcctt ccgccttcgc gtcacctttc taagaaattc ccagagggca 180

gcgcagacgg ggcgggctct gagactccgg gctccgcctc tttccgggaa ccgcccacta 240

cccaggactc cgacagaggg tgaaaaaaga taacttccgg tctcgcgatc gtctctaatc 300

tcgcgagaag agaaggcggc cgccatcggc cgaacggagg cggtggcgag ggagggggtg 360

tggccgggga gcgcgaagtc cccgggagta agggagaggg ggcggggtcg cgcgtcccgg 420

gcatacgcat gcgtgcacgc tgccggtcgg gctgggctga gaggggaggg ggcggcggcg 480

gccgaggcgg cgtcgttatt tccgtggtcc ggacagtgcg tggcggcgcg ggtgaccacg 540

ggagaagtag gcataatggt tatgaaagct tctgtagatg atgacgattc aggatgggag 600

ctcagtatgc cagaaaaaat ggagaaaagc aatacaaact gggtggacat tacccaagat 660

tttgaagaag cttgtcgaga attaaagttg ggagaactac ttcatgataa gctatttggt 720

ctttttgaag ccatgtctgc tattgaaatg atggatccca agatggatgc tggcatgatt 780

ggaaaccaag ttaatcgaaa agttctcaat tttgaacaag ctatcaagga tggcactatt 840

aaaattaaag atctcacctt gcctgaactg atagggatta tggatacatg tttttgctgt 900

ttgataacgt ggttagaagg ccattcactg gcacagacag tatttacgtg cctttacatt 960

cataatccag actttataga agatcctgct atgaaggctt ttgctctggg aatcttgaaa 1020

atctgtgaca ttgcaaggga aaaagtaaat aaagctgctg tttttgaaga ggaagatttt 1080

cagtcaatga cttatggatt taaaatggct aacagtgtga cagatcttcg agttacaggc 1140

atgctaaaag atgtggagga tgacatgcaa agaagagtaa agagtactcg aagtcgacaa 1200

ggagaagaaa gagatccaga agttgaacta gaacaccaac aatgtttagc agtattcagc 1260

agagtgaaat ttactcgtgt gttactgaca gtgcttatag cctttactaa gaaagagacc 1320

agtgctgttg cagaagctca aaaattgatg gttcaagcag cagatcttct ttctgccatt 1380

cataattcat tgcatcatgg catccaggcc cagaatgata ctacaaaagg agatcatcca 1440

attatgatgg gttttgaacc ccttgtgaac cagaggctac ttccacctac cttccctcga 1500

tatgcaaaaa taattaaaag ggaagaaatg gtgaactatt ttgcaagatt aatagataga 1560

ataaaaactg tctgtgaggt tgtgaattta acaaatttac attgtatcct ggatttttc 1620

tgtgaattta gtgaacagtc accatgtgtt ctttcaagat ctctgttaca aaccactttc 1680

ctggtggata acaaaaaggt ctttggaact catctcatgc aagacatggt gaaagatgca 1740

cttcggtctt ttgtcagatc ctccgagtgc tttcccccaa gtgctaccta tataataatc 1800

accaggctaa ggactgtatc gactcctttg ttactcactg tgttcggcca ttctgtagtc 1860

ttattcagat ccatggacat aacagggctc gacagagaga taagcttggt catattcttg 1920

aggaatttgc caccttgcag gatgagttta tgacatttta ttttaatagg cagagaaggt 1980

tgatgcagcg cttcacacca tgctgttgaa acaggaaccc caaaggcaac atttggcctg 2040

tttaggtacc tgggtccttt accataacct tcgcattatg atacagtacc ttctaagtgg 2100

ctttgaattg gaactctaca gtatgcacga gtactattac atatattggt atctctctga 2160

attcctttac gcatggttga tgtcaacatt gagtcgtgcc gatggctctc aaatggcaga 2220

ggaaaggata atggaagagc agcagaaagg ccgtagtagt aaaaaaacaa agaaaaaaaa 2280

gaaagttcgc ccattgagcc gagagatcac aatgagccaa gcatatcaga acatgtgtgc 2340

tggaatgttt aaaaccatgg tagcatttga catggacggc aaagtacgta aaccgaagtt 2400

tgagcttgat agtgaacaag ttcggtatga acacaggttt gctccattca acagtgtgat 2460

gaccccgccg ccagtgcact acttacagtt caaggaaatg tctgacctca ataaatatag 2520

ccctcctcct cagtctcctg aactgtatgt ggcagctagt aagcactttc aacaggcaaa 2580

aatgatattg gaaaatattc ctaacccgga ccatgaggtt aatagaattt taaaggttgc 2640

caaacccaac tttgtggtta tgaagttatt ggcaggagga cacaaaaagg aatctaaagt 2700

tcctcctgaa tttgatttct ctgctcataa atattttcct gttgtgaaac ttgtttgaga 2760

gagactgggg aggtggccat aaaggggcag agtcttcttt cagacccaac tcttagaggg 2820

cacatcacca ggctccacat cacgggaagt gagatggatt tcttgggtaa caactcatta 2880

taaggaatac ttttagtttg acagccttat atgacatgaa tgaaaactgc tgttttaaag 2940

tggtttatta tgttccatgg aagaaactgg tcttattgaa tgcattgatg aacgttatat 3000

ggttttatta cagatttaat cacaaatcat tttttatgaa tgattgagtg aaaatagtgt 3060

ttataaaggt taataaattt cttgacaaaa aaaaaa 3096

<210> 43
<211> 965
<212> DNA
<213> Homo sapien

<400> 43
gcgtggtcgc ggcgaggtct ttttttttt ttttttttt ttttttggga tgggaaaatt 60

ttattaaaat ggggaacact gtttaaatct tctggggcca tgaaacccca tcaggcagtc 120

taaaaaaacc atcggggagg tctgaggatc acttgacccc aaaattttga ggtctgtata 180

agctgggggt aaccggggct catagcgtgg ttcccgggtg tgaaatggtt acccgcctca 240

caaatcccac aacaacataa cggagacaag gagcctacgg tgacaaccac cctaggagca 300

gcccataata agaggagaac acaaacacac agacacatgg cgagcacaga aaaagaccag 360

aagacacaac gacggggaca cacgtgcgag gccacggcag cgcataaaag agaacgaggg 420

cgcaacgagc acgacgggga gaacaaacgc gaggagaaca ggcagaaaaa taggagcagg 480

ccactactcc ggatgaacca cccggcatca accataaaca caccactcag ccccaccccg 540

agacccgcta cagacaaagc caacaaccga cggctaaaac caccacacct tccacgcaca 600

aaaaaagcgg agcgcgaaaa taccaggtgg taaccaccaa cacagaaaaa catacgagcg 660

gaaaaacaca cgaccaggta aaaaagaaca attgtgtaag cgcaaaaacg gaccaacaaa 720

aaacgacgca gacaggcacc accggcaaaa aaaggccccg cagcatagca tgagggtaca 780

tcacacaaga cagactcagg acccacccag cgacagaagg cacaccaaaa aacgcgacac 840

ccacaagagc tcacacggtg gcaccaacaa ccccaacagg acacagatcc agaacaacca 900

aggcgggtgc cccccaagaa aacatccact agaggggact ccacaagaca cgaagccacc 960

gaccg 965

<210> 44
<211> 325
<212> DNA
<213> Homo sapien

<400> 44
aaaaaacgca gcttgttggc acaacacctg tagtcccaac tgtcttaaga ggccttgcgg 60

gcaggaggca ccactttgaa ccccccgggt gggtgtgggc ctgcccttga gctaatgatt 120

cgtgcccact tgcactccaa gccctggtgt tgaccgatgc aggaccctgt tctctctgac 180

accaggtctt ctctcggtgg tgttttgggg ctgcttaccc acaatttttt caccttggtt 240

ctcttctggt ccctaatact ggctcgaaac caacctttcc agttcttatt taaacccaaa 300

aaacccttgt tggtccaacc tggcc 325

<210> 45
<211> 333
<212> DNA
<213> Homo sapien

<400> 45
gatgactaat gggcgaatgg gccttagatg catgccgagc ggcgccagtg tgatggatgc 60

gggccgccgg gcaggtactg ctgattttca gtctaaggac atatatctct tatatcatat 120

tgcctcttaa aaggtaaaga aaggcaggtt ggacccatga catatcttct aggccacagc 180

tctgaacaca ttgcaagaga aatattcaag caaagtgaaa ggaaagcagc acattttcag 240

catcttaata gtgaagctat catactgaag gaaaccatat gagaaaggga tatagaaagg 300

gcaccccttc tcttcatttc cctctaacac tgg 333

<210> 46
<211> 273
<212> DNA
<213> Homo sapien

<400> 46
cggccgaggt gtagggtgtg tggtgtgtgt ttagggtgtg gtgtgtggtg tgtgatgtgt 60

gtgtggtgtg tgtggtatgt agtatatgtg gtatgtggtg tgtgtcgtgt gtgtggtgga 120

tacacaactc tatactaaaa gccaatgagt tgtttactta aagtgggtga actttatgct 180

atacaaatta tatctcaata cagatttctt taagtcttca ggaagccctc tggtaaagaa 240

gtcagcctaa cccagccctg cactcatctg acc 273

<210> 47
<211> 1526
<212> DNA
<213> Homo sapien

<400> 47
ttttttttt atgattaagg aattctgttc attaaaagag atcaacaatc attacatatt 60

ttatgcttgt atcaaaatat tacatgtacc tcataaatat atacaacaat tatgtattgt 120

tcttctatta catatagcag tttagaagtc agactgttac cactgcagat aacgtttgat 180

tttcagcatt tctataaaat ttccataaaa attaaaaatt ttcttaaaac aaattaaaga 240

tatcaataag taaaaaagta tatatttgca atgcatatat ttgacaaaag attcatatcc 300

agaatacata aagagccctt acaaatcaat gacaaaagac atctaaaaga caaacaaaac 360

aagatgtaca atggccagtc aacatatatt gaaaagattc tcaatttcat tagtcatcag 420

agaaatgcaa aaggaaacca taatgagagg tcaccacatg atcaccacat tggctaaaat 480

aaaaaatacc aaaatgccaa gtgttggtga gaatgtaggg aaactggaac tcgtgtacac 540

tgctggtggg aatgcaaaat agtgcacctg ctttggaaaa gagtctggga gttcctctaa 600

aagctcaatg tagaattacc atatgaccca gcaattccac tcctctgtat agacccaaga 660

gaactgaaaa catatggtca aatacaactt gctcatgaat gtttataatg acgttattta 720

tgatagccaa aaagtggaaa caacccaaat gtccatcagt gcatacatgc aacaatgtgg 780

atgaaccttg aaaacattaa gttaaatgaa agaagctggt cgcacaaaga tcacacagta 840

aatgagtcca tctgtatgaa aagtcccaag aataggccaa tctatagagg cagaaggtaa 900

attagtggtt gtcaggggct aggaaggaag tggatgggaa atggctgcaa acagcatgag 960

gtgttttggg tggtgatgga aacattctgc agtgacattg tggtgatgga tacacaactc 1020

tatactaaaa gccaatgagt tgtttactta aagtgggtga actttatgct atacaaatta 1080

tatctcaata cagatttctt taagtcttca ggaagccctc tggtaaagaa gtcagcctaa 1140

cccagccctg aactcatctg accaccaaag cttttcctca cattggcacc ctgagaaact 1200

ggtattctga agaacgcgct ttaggaaaaa ctgctttaga caacaggaat ttggtaagaa 1260

gaactttgtt tctgtgaaca catatttgca tgtcagggta catccttttg tatattattt 1320

atatttagtg tgtctatgtc ttgtcttctt ggtagcttta caagaatttc gaggagagaa 1380

agtatgattt tgtctctttg aattcctact tctcaccacc cataatgtgg tgcacacata 1440

aatatctgta aatatgcagt tagaactttg catcactaat gagttaatta aactattcaa 1500

caaagccaaa aatacatatc atggtc 1526

<210> 48
<211> 962
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (53)..(662)
<223> a, c, g or t

<400> 48
gcccatggat actaacttct gcagttatcc attcagaaaa ttttcagaca tgnnnnnnnn 60

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnatacttc cgattcctcg catcaactga 420

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 660

nnaacagaga tgagcgacac acacagaaaa ggaactaata caaggaatac aaaacgcggg 720

ttaccacacc atgcgaaaga cacccatcga agcagcaaac caatagtccg aaaccgtggc 780

aggaataacg gaataactag acacgcatat tatcccaaaa gagaaagcgt agcagcgtaa 840

acaaaaacac acagcaggaa ggaatcagac gaaagaagga cgaacgacca gagcgagggg 900

aaaaaccccg aaaaacgacg agctaacgga aaacgccgaa acaacggaga agaatatacg 960

ga 962

<210> 49
<211> 1757
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (609)..(609)
<223> a, c, g or t

<400> 49
gacccggccg cccgggcggg taaaaactag aaggtttctc ttccacttga tgttgtgccc 60

cttcactcat attgattgct gagagacgat atgaggggtc atcatgtgaa cactctgaac 120

gctaaatgta atactaacgt gctcatcgag cagcggcagg gattagcgat actctcttca 180

ctgaccctga tgttcttgag tccttggggt ctactccatg gtatgtcccg ttaagtcttg 240

tgagccagta catctgttac ttcgtgtccc tggaagcgtc ttgaactatc attttccccg 300

tggaaaacca gtggcaaggc cgtccaaagt cgagtgtcgg tattcacaag agcgtcgagt 360

gtattggagc atattggctc agtcacaaac attcaagtcg tccctccctt cccatggtcg 420

accatgcata gagagctggc gtactagcgg cgtccatgct agaaaacctc gacgtagacc 480

tatacatcct ctatgtgcaa aacgatctat aaaggaccgt aagcttcgcc caaaggtttc 540

tacatcggta gagccacctc ccgggcataa cgggtgggtt ccctcctcaa gcccccttaa 600

atctatacnc ctcgccgtgt ttggttcggt gccccgcccc gttgaggcaa cataccctaa 660

cattggcaag attgcgggct ttgcaaaagg gcatcgctag taaggggcc ttgtgagcct 720

aaaaaacccc gccttttag cccctacgta caccagtacc gaatttccct ctcttggggt 780

cccgcccgag ggaacacaaa tggtcacctt tcctttcagt gccatactcc accttcccc 840

cgcgcgtagc ctgaccttac ttaattggca atgagggtca aaatccaacg ttcgcaataa 900

ggtggccccc gaaatatgtc agggcattcc ctagtttgtt ctattaacgg gccccaatgg 960

taacgtcggt gcgccgcaaa gtctccgatc acggggagaa ttatgctgtt cagacacaca 1020

acagagaaac gcgggggcat tatcacttaa tgttgtatct ctacgcaagg gcgtaactcc 1080

cccccgcgtc accctatcac tcaccgttat gtggcgattg tccaatggaa gcactcccaa 1140

gggacgcggt tgcgcatttt ggccttttgg aagagtttgt ggcccccccc cattggggct 1200

tgtttaaacc ctttaggtgc ttggcacggt gtgggtctgg ccccccctcgg ggtccccttt 1260

tggaagggtt cccaaggtta ggtctcttct cgttccaatg gacttttagg cggccaaagg 1320

ggttgatcca caacgttgcg accgttttag cattagagga aacacgtgat taaggggaaa 1380

gcggaagaaa acaaagggtt taccccttcac acctgtgcgg gggtacaagc taaaaggtcc 1440

acaacggttg tgtcacagaa acccccccgca gatggttttg cccttcacaa ccccttccga 1500

aaaaatggtg gatggggcgc cccttggtgc ttcacccccct tgtgggaaag gttcacaccg 1560

gtgggttgag acacctctgc gaaacaaggg ggccctatgg ggagcgagga aaaacatcac 1620

actgtgtgtg cccattcaac ggctttgtgt ttcccgctgg tgtaacattg aggtgatcga 1680

gatcaacaca cgggctccac atatccacac agacgacaac acgacacaca acgacacaga 1740

cacagacgaa cagacgt 1757

<210> 50
<211> 1670
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (293)..(293)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (1029)..(1029)
<223> a, c, g or t

<400> 50

gcgggccgcc gcggggcagg tcttttattt aaatagccat gatccatgat agggatgcta 60

gatctatgat aagaaaagca cgacacagtc cactggaact gcacatagtc ctcaaaaggc 120

agggatcatg catccgtcca ggcatactca tgtatccaag ctttatcatc attcaactac 180

acagcgtgct aatttgaagc ttatcttaga ttccgataat cccactggca acctagcttg 240

agtaatgagg cgcgcgcgag gagcgtaagg ccgttcgtgg cgttgtaatt gtntctcttg 300

aagcgcccac tcgcgtcgta tatttgcggc gcagtaatct ggtgcgtcca ctcatgtccc 360

agaggtcttt cacgatctta gctgagtgac tgggagagac tgtgtgtcaa ccacacgcat 420

gcgcgcgagt tgtgttcttg ggctctcctt ttgcagtata cactagttcg atggcacagg 480

gaccgatatg cgtgtgcgcg cattgagttc cgtgcatata tgggcttttc tatacaacat 540

tatctctgcg cgttcaggtc actcgagaac tcccaataga gggtctagag ttggtggatt 600

ctatccatcg atatttacct aagtacattg tgttagcggc cccactttct cctcggtcag 660

ttgtcttctc ccacgaactc gggcggatgg cacccattat ttctggcccc agagagcctg 720

tggcagcttt agtccatggg tacagttagt acatggctcg tgtctcgtgg tgtattgccc 780

acatttgctg tgaggtgcgc gcctttcgac cgcgagtggg cctatctact tgagctacgg 840

gggggagaag ggcgagggtt cccgggggct ttattctccc cccgttagag cccgttggga 900

agggcccggg cctgtggctt ttcccacttg ggggtgtgtg ggggtccact tttttgggcc 960

gtgtttccac gacatgccca cacgtggttg tgccccacgc ggacggtgaa acccaattcc 1020

ctcttggtna cccgagacac agccgttttc ctcttgggac catttgggaa agggcccccc 1080

agcaatttgg gtcccccccg gggaaaagga aagggctttc tcgatctcaa aagggccttg 1140

aggagcaggt acacctccat tggggtcaca aaaggcgtgt gtttcgcggt ggtggttgcc 1200

aaacattgtg ggttattccg ggcttcagca caatttcccc acacgtctac cacgaggagc 1260

accaaaaatg agggagcagg acgaagagga caaggggggag agagcgcagg aggcgagagg 1320

cggggcagac caggagcagc agacagcaca aacggaagac ggaagaggaa gtaacaccgg 1380

acgagaacag ggcaaggcag ggacccaagg cggacaaagc acccgcgaac tcgaacgagc 1440

gagcagggga gcgcgcgcag gatcgagacg ccacgagaag gagaaggcat cgaggagtag 1500

gaggcgccga ggcagaacac acgccacccg agaccagaac agcacgctca gtggacgcga 1560

cgagacacgc ggacagacga gacgaaacag aaggagtcct cgaacgcgga caggccagca 1620

aggccagaga gcgcagcgcg caccaaagaa gcgaagaaaa gggacacgta 1670

<210> 51
<211> 148

<212> DNA
<213> Homo sapien

<400> 51
ggcgagaaag tgatactcaa tataggcgac tggccttata atcatgtcga gccggctgca 60

gtgttgaatg gatagcgtgg tcgcgacgag gtacttcctg ggtgggccaa gccacccccag 120

agagttatgt ttaccgagga ccctgaag 148

<210> 52
<211> 393
<212> DNA
<213> Homo sapien

<400> 52
gctttttttt ttttttttt ttttggccag tgctttctac tttattaaac atcagagagc 60

ccaaatagaa tgtccccggg ggagggaggc acttaagagg caccactaga ggggagagga 120

gaaagagggc acccctgggg aaagaagaaa tccaccaccc acaagaagac accaactctc 180

tccacaaaaa gagggctcca cacaatttga ttctcctaag gggaggacgc aggcgcaggg 240

ctccacggcc ttcaaaattt gtgggtgata taacgcgttc gaggatgtag aagggacccc 300

caagcctggg cggttaaact cagtgggctc aatagccgtg tttcccgtgg tggtgaaatt 360

gggttactcc ggctcaccaa ttccccaaca aat 393

<210> 53
<211> 574
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (156)..(156)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (176)..(176)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (215)..(215)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (226)..(226)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (234)..(234)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (237)..(238)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (272)..(272)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (277)..(277)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (299)..(300)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (306)..(306)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (343)..(344)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (383)..(383)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (428)..(428)
<223> a, c, g or t

<400> 53
tacacggccg agcatgttca cgcacgtgga tccgagagcg ccgctgcact tcagtcactg 60

ttcttacgcg ccccgtggtg atggacacgt gccgagcgtg ctgcgagctc gagctctgga 120

attgacgctg cggaggacaa catacgaact aggcantgtg aacgactgcg ctaaangtcg 180

tacttgttgg gctaagacgg gtgcgacttg acacngcgtc tcaacntcga cgancgnnta 240

gtgcatcctg cgctcagcgg ggggttgccg antggantag cactctcacc ctttaatann 300

gcttgngctg ctaatgtcac tggctgcact agcgttgtgt tcnncttgtg ctgaacactg 360

tgtttattcc tgcatcgagc aanttcgcat cgatcaacat aaggaaagcc actgcgctcc 420

actccggncg tcgacgaagt gtcacagcga gcagacgcag tacgaacgcc acaagcgagc 480

ccacccccgc agacgcgcgc ccgacacacc gacaccgcgc gccagccgcc acgcgccaac 540

cgcgcccacc acacccacac caccgcacca ggcc 574

<210> 54
<211> 1332
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (389)..(389)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (646)..(646)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (989)..(989)
<223> a, c, g or t

<400> 54
gagcggccgc ccgggcaggt acaatcttat ttattgaaca tcttgaggtg ggcatgggcg 60

agagggagga aagcagctac ttcggtaaac gagtttctac aagaactatg tgctcagtaa 120

cccgggtgct ccagttgtac gtgtagtgaa acttcgactt ttccacaaca ttggcaggca 180

cgaccatttt ccgtgtcgca tgggtggact atatggatca gcagtggagc tgacctgtcg 240

agcgtctagc actgaccttg actgggactc acctttcacg atcccacgtt ttgccattat 300

[redacted]

aaatgggtat agtatattgc aatcctcant tcattcagtc aatacaggcg taatttattg 420

ttgtacaagt tgtggagtca cgttgtaaaa gagtgctact ctaagagaag cagaaatacg 480

ggtacgctga cttcacggtg ggtgttgatg cctcgccctt cgagcgactc tgttggcggg 540

<211> 468
<212> DNA
<213> Homo sapien

<400> 56
ggtgtatgtc tataggccct gttatctaat gctgctcagc cggcgcggta tgtgatggat 60

gtggcgcggc cgaggtactc cttcaacaag ggatcgaccc tagctactca ggaggctgag 120

gtggaataat tgtttgaggc caggagttcc agatcagccc gggcaacatc atgcgacccc 180

atctctaaaa acatcttttt aaaaatgagc caggtgtggt agcatgcacc cgtagtctca 240

gctactcagg agcctgaggc aggaggaagg tttcaacata ggagatcgag gctgctgtga 300

gctatgatcg tgctactgca ctccagcctg ggtgacacag caagttcctg tttccaaaca 360

acaacaagaa aacaaaacaa aaaaaagaaa aaaaaaaaaa aaaaaaggtt ggggtattgg 420

gcaagttccg gtggtggatt tttttcccgg ccatccccaa tttgaaac 468

<210> 57
<211> 499
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (243)..(243)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (258)..(258)
<223> a, c, g or t

<400> 57
ccgcccgggc tggtacacga gcgaatggct agatgtttac tcgctctcac tgctgcgaga 60

ccatcagcct gctcaatcga cttgggtagg ccgcgacgtg acaacaacct gaacggccag 120

acaagcccgc aagtcggaat cgatcttcca tggctacggg ccttgtggca cgagcacgct 180

ctagtgctac acgcgagcaa tcttcagcac gctagccact ggctagccac cgagagcacc 240

tgntctccgg ggagcagnca tttgaactcg taggcgagca acgtgagcac tcatcgagag 300

aacgggtcag ccgttgggcg ctaggtcact ggctcgatag gctgctcctc ctgttgctga 360

atagtgcttc tccgcttcac aggttccagc tacaacgaga cgagcagcct ttgaccaggc 420

aggtcaggct gacctggttc ttggtcagct catcccggga tgggggcagg gtgtacctcg 480

gccgcgacca cgctaagcc 499

<210> 58
<211> 424
<212> DNA
<213> Homo sapien

<400> 58
ccgcccggcc aggtactctt agtacagatg gggtctcacc atgttgcgac cagactggtc 60

tctaacattg tgacctctga agctgatcca acctgccctg cggcgtcccg aagagtgctc 120

gggattacta gcgcaacgag ccactatgcc tggacctcta ttgttcatgt acataccatg 180

ttcttacaga tagtgaaaat aggtcagata tcttagaaat aggtattccg tgttcgtaag 240

ttccgactgt ggatatgaat gcatatcttg gtgtattgtc tgcttgctca gataaatgat 300

tcatcgcaaa ccacgacaac ttggtccaat ggtgacgttg ttcatctttg actttaagac 360

aagatgcatg catagttcat atcactagag tccctttcaa gaacagaggc ctgctcgtta 420

catg 424

<210> 59
<211> 1264
<212> DNA
<213> Homo sapien

<400> 59
cctgaaaggt ttttcccgtt cgtgcacacc tcctttacca ctagcccttt cgggttttac 60

acccgtggcc gaccagggcc attgaaagcc cgttccatac gtaaaatagg agaggacctc 120

aattgtttgt tttgagcaga ctttgcccgt cccagctgac catacgtgaa ctcggactta 180

cgcaacggcc cttccaagct caactactcc cacccaggct gggactacag gcacatgcca 240

ccacattcgc ctaatttttgt attttctgta gagacagagt ttcaccatgt tgcccaccct 300

ggtcttggaa ctcctgggct caagggatct gctggccttg gcctctcaaa gtgctggggt 360

tataggcatg agcctataac cctcaaatat cttaagaaaa gtaactgact gcagttgaaa 420

acaggtaatt gaaattgtgg taagtgaaac catggataaa gcgggactac tgtacatgct 480

cattaaaaaa aattaagggc caggcatggt ggccttacac ctgtagtcct agctactcag 540

gatgtctgag tcatggaatg actgcctgat tcccagtgtt gagctcgatg gcgtactatg 600

atctgtgatc acacaccact gcactccagc ctgcgtacca caagatcctg tctcaaaaaa 660

tatataaagt aaaaagagtg attttattta tttatgaaac agggtctcac tctgtcgccc 720

aggctggagt gcaatggcat gatcttggct cactgcagtc tccgcctctt gggttcaagc 780

gattctcttg cctcagcctc ctgagtagct gggactacag gcactcgcca ccatgcccag 840

cgtaattttt ttgtattttt agtagagatg gggtctcacc atgttgcgac caggctggtc 900

tctaacattg tgacctctga agctgatcca acctgccctg cggcgtcccg aagagtgctc 960

gggattacta gcgcaacgag ccactatgcc tggacctcta ttgttcatgt acataccatg 1020

ttcttacaga tagtgaaaat aggtcagata tcttagaaat aggtattccg tgttcgtaag 1080

ttccgactgt ggatatgaat gcatatcttg gtgtattgtc tgcttgctca gataaatgat 1140

tcatcgcaaa ccacgacaac ttggtccaat ggtgacgttg ttcatctttg actttaagac 1200

aagatgcatg catagttcat atcactagag tccctttcaa gaacagaggc ctgctcgtta 1260

catg 1264

<210> 60
<211> 1512
<212> DNA
<213> Homo sapien

<400> 60
gtggtcgcgg ccgagcgtca catttccaat cttaatagac gcatagccag acttctgctt 60

ctgatgactg agctacaggc tacagtgagc taggctccca accggttctc aacattctgt 120

attgttggta taattattct cccagcactt ctatactatt gtctgcccgt agtgcctcgc 180

taagagagca catgctaggc tcagttatgc tcgaagcgag acatctagtg tcttcgacgc 240

agcggctata tagctggcta tcatcaaagt cccaggctct cgagcccaag aaggcctctg 300

ggcgctacac tccaatggat cgactatgca tgctcctcgc ctgctgttag aatagttggt 360

ctactccagc tccacaactc tcacacacaa caactacgga aggcaaggta cactcgctgc 420

gtccagcaga ccactgccgc catttaacgc gcaggccgag cctcaccacg acatgcctga 480

catcccccat agtcccaact tccatgctgc tactgacgct ctccaccttа ttgtcccttg 540

ccaacactct aagccacttt tcctcgttca tcgccccccc aaacaacaca cacgcacatt 600

gttcgcctca ctcgcacagc gctcttgggc ccgctaaacg tcccattgcg ttccattaag 660

gccctcggtc ttcccatgct tgctgcaccc ttctggcttc tattctcgga cttccagcta 720

agtttcccta tccagccaca tcataccact cagtcatgca aatgccactc acccccgtca 780

ctctgtctcc cgccccatcc ctcccctctc cacccgtcgt ctccgtctca ccctcgcccg 840

gcgcgccacc tcctcccact ccgccaccct tctaccccgc cctcgcctac ctctctacca 900

gcactcccat ctctatcccc cctctcttcc atcccacacc acccccccтc caccacagcc 960

gccattcaac tgccccccac tccacaccac ctccgcccca cacacaacta ctcacccatc 1020

cgctccagcc actccacccc ctcacctcat aacaccccta gacccacacc cacgcccccc 1080

ccaccacgca tccactacac caccatatcc cccctcaaca ccacctcccc ccctctacac 1140

tcaaccctct cttccccacc ccctctccac caatacaacc cctctcaata ctcatacacc 1200

atcattcaaa ctgctactac ccacccccaa ctctcacaca cacccatgag aaccaacaat 1260

catcactcta tactctaccc cccatcgctc tctccaccac ccccacgcac ccgccacacc 1320

ccacccccctc atcaccgcca ccacctccta ctctacttac tccctccata cacccggcct 1380

cccactcctc tacgtcccca ctcctcctca accatataca ctcctcccgc ctactctctc 1440

ccgataactc ccaccatatc ctctctatca ccccaactcc ccccctccca ctaccacctc 1500

accacacaac ac 1512

<210> 61
<211> 775
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (12)..(12)
<223> a, c, g or t

<400> 61
cgtggtcgcg gncgaggtac ataccgtctt ttttttttt ttttttttt ggaaacagtt 60

tcgttatgtt tgcccaggtt ggattgactt ggcgcaatct cggttcatta gaacctccac 120

ctcccgggtt acaccccatt ctccgtgcct aagccccccg aaatagcgtg ggaataacgg 180

gcccccgcaa accacgaccc ggttaaattt tggaaatatc tagttagaag acacgggttt 240

tcccccgttg tttgcccgag ggatggctct cgaatcctcc ttgacctttg tgaactccca 300

cccacctagg ccttccccaa agttgctggg atacaacgag gcgtgaacca ttgccccccgg 360

ccaaattcac agttccttat caaagaatat accccagatt aaaatctctg ttgattgata 420

accgataatc cccaatatta gtgtaaaaat tttacggaaa agtgttatcc taaatagacc 480

tcttaggcca aaataccagg tctgtatgag aggccatctg atgccctcaa tctgtccagt 540

acatctccca gaagacctgt aaaaatatac ccctttttg gtggggcata tgaacttttt 600

caacgggagt agaatctcaa tgtgtagaac cagatgtccc tgaatggaaa atttggattc 660

ctaaaaagtg tgtcccttcc taattggctg tccctaattg gataattaaa tcctgtatta 720

tgaaaatctt gggcaaaacc tacagtttgc atattccatt accccatggt agttc 775

<210> 62
<211> 918
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (505)..(505)
<223> a, c, g or t

<400> 62
cgagtttttt ttttttttt ttttttttaa aaaggaaaac ccggtaatga ttgtcggggt 60

tagagggata ggaggaaaat gggggatagg cgtgttttga ccattgaggg gtgttttccc 120

tcggtggtga attagagggt ttaatgtgtt gtgtcttaag tggtgggtgg gtgtgagtgt 180

gacgccccat tgtgtgtgtg tgtggtggtt aaatatatgt gtataagagt gggagtataa 240

taggcgcgtg gtgcacatag atctagtgtg tgtgaagtct ccgtggtgaa gtataggaaa 300

cagagtgtat ctttgtgtgt atctcacgag aaacagatgt gtgtgtgtgt taccatagac 360

acacacagag agagagtttc tcctccccga gatatagcgt gtatacaata gagtgagggg 420

gggagtataa agggcgcgaa gaggttataa gcgcgagaag agcccttgtg tggctataag 480

aaagagttct ctttctctac aaaanagagc ggctttattt attagatgtg tgaggcgagt 540

tattagaaaa gtcttttgtg aaaaagtgct ccctctgtta gagagagaga gagatattac 600

tacgtgtatg gtgcgcgcac gcgttgttgt gaaagatggt tgcgcgctat cgcggaaaag 660

gaatgtgggc acgattgttg atggccggtg ggggccccac gacacatatg agttatacat 720

gatgaggaga gagaatgtgt ttaacaggtc ctccccgggg gggggggca gcgagaatta 780

ttatttgtag aacaatatgt gatagctgtt gtgcgccccc gccggtggtg ttaaaaaacg 840

cctctaggtg gggcggaaat aacacctccg agtgggggc tccacaatag gcgcgttgtg 900

ttccccgcgg tggggtgg 918

<210> 63
<211> 807
<212> DNA
<213> Homo sapien

<400> 63
gtcgcggcga ggtacaaaaa ttagctgggt gttggggcac gtgcctgtgg tcccagctac 60

tgggaggctg agccaggaga atcagttgaa ccaggagtca gaggttgcag tgagccccga 120

gatcgcgcca cagcactcca gcctggcaac acagcgagac tcccatcgga actaaacata 180

tataaaacaa aaaaaaaaag acgctggtgg cggttacctc gtgtggccat ggctgtgttt 240

cccgtggtgt gtggaaaatg gtttctctcc cgctccacaa aatccccact cacaaacttt 300

acgaagcaaa tgtccatgca caaatactga atctccaaat cgttatacat attttcgtga 360

tactgatacc tccaattaag gaacatgctt acacacggtt acagcattgc gaagtacgtg 420

aaatacttct cagagaacac gacggtagac ggcacgtaac acgagaaagc atcagagaga 480

gcgcctagtt cctcgactag acttaccgac tactgcctag gatatcacga caggttccca 540

gacatagggt actcgcacgg aacctggtag atggcactag gaagaccatt gaaaagagct 600

taattagaat aactataaac tacacccact attgaaaacg ttcaatgtag ccccagcgat 660

cgatgacaac ggcggaaaga tgaacagtaa agcacacgga ggcttacatt tcctagcctt 720

gacttattta acctggacta taagaataaa aacaaaggca ggagagcagg caacaagaaa 780

tataccataa agcgagctag cgcccct 807

<210> 64
<211> 513
<212> DNA
<213> Homo sapien

<400> 64
gggatgatga tcactatagg ggcaatggtg catctagatg catgctcgag cggcgcagtt 60

gtgatggatc caagaccagc ctggccaaga tggtgaaacc ccatctctac taaaaaatac 120

aaagagggag cttggcgtgg tagtgcgcac ctgtaatccc agctactcgg gaggctgagg 180

cagacaattg cttgaacccg agagacggag ggtgcagtga gccgagatcg tgctactgca 240

ctccagcctg ggcaacagag caagacgtcc gtctcaaaaa agaaagaaaa aaaaaaaaag 300

ctgggggcgt aatctcatgt ggctcattag ccgtgtttcc cgtggtgggt gagacattgg 360

cttattccgg cttcgacaat tctccaccac cagaacatta cccgcagacc agggggtgtg 420

ttcataggcg acgaaagagt aggagtagcc tgcatggtca tgccgatgcg atgaacatcg 480

ttataggcag atcacgtaca agtgacgtgt acc 513

<210> 65
<211> 432
<212> DNA
<213> Homo sapien

<400> 65
acgtatccgt cgcatcaact gaactcgctg acgctcggat cgctgtcggc gtgcgagacg 60

agcgagatat cagactcaca tcaaagagca gagtaaatac gagtattatc cacgagaatc 120

acgggagaat aagcgcagga caagaaacca ctggtggaag caaagaaggc acagacaaga 180

aaggcaccag ggaaaacacg cgaagacaaa ggcccggcgc taggccttgc gcgatacaga 240

accgcaatca ggacatcccg cacacacacg catgaacaca gccaatcaac ccaacgaaaa 300

ttcgaaacgc agtccaagat ccgagacgga tggcggacga cccccgcaca ggagactaag 360

tagaaagcaa tacacaaggc agttggaccc cccgtggaag cgtccacatc atgagagcgt 420

actccactgt ac 432

<210> 66
<211> 457
<212> DNA
<213> Homo sapien

<400> 66
gcgtggtcgc ggccgaggta cttatacccc ctaaatatat aaaacatttt taaaagaaaa 60

aaaggaagaa actattcata catgcaacaa cttggatgga tttcaaggga attatgctga 120

atgaaaaaag atcagcctcg taagattaca ttctgtatga ttccattcat acaacattct 180

tgaaatgaca aaattacaga gatggaggac agaacagtgg tagccacagg ttggggtgag 240

ggtataagaa agggatgtgg ctgcggttgt aaaagggcag tgcaagggat ccatgtgaca 300

gaactgttct gtctcttgtg atggtggtca catgaatcta cacatgtgat aatattgcat 360

agaattaaat acacatacac gaaaaaagtt caagcagttg agcacaaata ttttaattgt 420

ctaaaatgac attttcttta agagttatct acagttc 457

<210> 67
<211> 2593
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (2340)..(2340)
<223> a, c, g or t

<400> 67
ctattatgtt tcccaatttg tccaggtcct tcctccgtgt gtaagtagcc cgagaaggct 60

tccacattcg gcgctttcta ggctccccgc ccggttttca gcccatgtcc tcccagttgc 120

cgccgcaggt gccaatggtg tgacagttaa cccgacagaa ctacttttat gcctcaggag 180

aggaagacac aaggagtcaa aaggggaaa aaaaaagttt gggttcatag tagcaggaac 240

attaacagaa tagcctgaga ttttaacagc ataactcatt ccctcttcca cctttgtact 300

ttatccaggt caacacatca gggttctcta acgattccag tattctgttt ctttactgta 360

agatacatgt aattcttgcc actgtgatta aacaagccct gtaatagtca gcagggttaa 420

aaagagatta cggaaaggat aaactcctac ctactttctt gggagatgtg ggaaagattt 480

caagtcacag cattttcat gactgtttat aaacaatggt catttatatc cacactttct 540

cttatttaca ttagttttgg cccttaggca actcatactc ctacagtgat tattggcttt 600

gctttcataa catgtatttt taagtattta ctctcttaat ggccctcgat gtctatttta 660

tacatcatat ctcttaattc tctagatgga acactgaagg acaggaatta agtaagtgac 720

tggccatgca agggttggaa attttactta tttttccttg gtagaagtta tgttaaaaat 780

tcaagcaacc acatatctaa cagaggaatt ttatctagga tatataaaaa acctctcaaa 840

actcaatagt aaaaagaaca aatgacctaa atagaaaata gacaaaagac atgaagacat 900

ttcaccgaag aggatacata gatggcaaat tagcacacaa aaagatactc aacatcatta 960

gccattggaa atgcaaatta aaaccacatg tggtatcatt acacacatct atgtgaatgg 1020

ttaagataaa aaatagtagt aataccaaat gctggtgagg atgtgaagaa actggatcaa 1080

tcatacattg ctgtctgaat tgtatgagtg gctgtatgta aaaagtagag ccactctgga 1140

aaaagagtag ggtagtttct tacaaaaata tatgtgttta ccatacaacc caacagttgc 1200

ccttttgagc atttatccca gaaaatgaaa atgtatgttc acataaaaac ctgtacatga 1260

atgttcacag cagctttatt agggcaaaaa actgaaaaca actcttatgt cctttagtgg 1320

gtgaatggtt aagcaaactg tggtacatcc ataccatgga atactactca gcaatcaaaa 1380

ggaactgccc ccacttcacc acgatgcaat atatgcatgt aagaaatctg tacttatacc 1440

ccctaaatat ataaaacatt tttaaaagaa aaaaaggaag aaactattca tacatgcaac 1500

aacttggatg gatttcaagg gaattatgct gaatgaaaaa agatcagcct cgtaagatta 1560

cattctgtat gattccattc atacaacatt cttgaaatga caaaattaca gagatggagg 1620

agcagaacag tggtagccac aggttggggt gagggtataa gaaagggatg tggctgcggc 1680

tgtaaaaggg cagtgcaagg gatccatgtg acagaactgt tctgtctctt gtgatggtgg 1740

tcacatgaat ctacacatgt gataatattg catagaatta aatacacata cacgaaaaaa 1800

gttcaagcag ttgagcacaa atattttaat tgtctaaaat gacattttct ttaagagtta 1860

tctacagttc aaagcccact tttatgaggt gtcacatcca tcaccatttt aagagatata 1920

aaatcatgaa aagatatcac cagaagctat gtaaacattt cagctaaggg taaagagaaa 1980

gttaagggtg ttttcacaag gaaattgaaa gagggcaatc caaatgaagt caacatggtc 2040

acacaaaaat cttggtaaaa gaactagaat ggaagcccaa gctgctgagc aagtgggaga 2100

agaaaagaaa acatagtcca aacagatcac acaagggaac ccaggacaaa tgctgacttt 2160

ggcattatct aggtaacccc tatttgtcgt catacgcgac tctaataatg gacctatagt 2220

tgcaaagcca gtcatagtcc taccaaattc aagagaggtc cctcattgac tcggggatgt 2280

agtgtggacc ccatgtcccc cacaccaaac cagatcattt gtggtaagaa agcaccaccn 2340

gcttttgtgg cactgctgta agaagcaata ggccggcacc caccagagat gttcttgtgc 2400

ttgtgaacaa gaccgagaaa aattgccttt tcatcaagta ataaatcctg gccttaaaaa 2460

acgctccagt gattacccac tgggggataa ccaggcgacc accatcatgg accccatttt 2520

ttgtcccaag attggggatc tattaataac aatttttctt ttttttaat ggggcaacac 2580

gtaaccaaaa ttg 2593

<210> 68
<211> 1253
<212> DNA
<213> Homo sapien

<400> 68
tgcggccggc cgggcaggtc ttgcctggat gaggccagcg gacacatgaa gagaagccca 60

gtatctcatt taatcttaag agactctcta tgtcaaggat tcccgtgtgg gggctgaaaa 120

tgtacagtga gataaaatta tgaacggcca cttagtcatc acgtccattc gtgcttgctc 180

caatgtttcc atgggctgga cgcgtctctc aagcagagag gctaatctga ctcttatgct 240

aggaagactg atggctgctg ggactaagga cccagaacag ttccatgaga tgaggcgacg 300

acgattacga tgacccctcc gctagtgccc agatggtgac ccactttcgc gtctgctcaa 360

tgtgccagtg cttcgaagtg gatccagctt gctttctgaa ttagtgagtt cctggagcta 420

acatgatggc cataatcgga ttctttcacc gctcttggag cagcaaagct catggactag 480

gaacactggc taagaagcga aagcacacaa atgagaacgc ggaaagatcg aaaaaggcag 540

gtgcagacgt atttgaagga aaagccctga aaagtaatgc cgtgtacatc cgacagcttg 600

gactgttcct gtgtgtgcaa agcacacgta agaaatgtaa ggcagagaag atctcgttac 660

gcatgggtca gtccattttt atggaacccc tttttgcgtg gggacagggt gtgggatgag 720

cggaaacctt ttaatgcatg gttcccatag tcaaacttca cccgccttga tatgggcaac 780

ttttggagcc cagtacaaga aaacagttgc ccgtcaagaa gaagcatcgg tatcgagggt 840

aagcccttag ggggttgggc ccctagttga atgtcaattg ggttgaattt cacgccaaga 900

atggttgctc gagatatggt atactttgtt ccaattctgt ggacttggag aacccatgca 960

attgactatg ctaaaaggag agaaaccaac acgtgggtgc acacccccaa aattccggcg 1020

ttgaagagaa ggcactccag cgggacaatt tcggcaacaa attgggggggg ccttttttacc 1080

caagggtgca aagttggaaa ggaaaagccg tcccttcccc taacatccca tgagcaattt 1140

tgcgctggag tatacccaat taatacaacc caaaggacaa ttatccctcc aaggggtctt 1200

ttaccctccc tttccccttt acctggagaa tttaccttct ttgtgatgtg gcc 1253

<210> 69
<211> 454
<212> DNA
<213> Homo sapien

<400> 69
tggtcgcggc cgaggtactt atacccccta aatatataaa acatttttaa aagaaaacaa 60

ggaagaaact attcatacat gcaacaactt ggatggattt caagggaatt atgctgaatg 120

aaaaaagatc agcctcgtaa gattacattc tgtatgattc cattcataca acattcttga 180

aatgacaaaa ttacagagat ggaggacaga acagtggtag ccgcaggttg gggtgagggt 240

ataagaaagg gatgtggctg cggttgtaaa agggcagtgc aagggatcca tgtgacagaa 300

ctgttctgtc tcttgtgatg gtggtcacat gaatctacac atgtgataat attgcataga 360

attaaataca catacacgaa aaaagttcaa gcagttgagc acaaatattt taattgtcta 420

aaatgacatt ttctttaaga gttatctaca gttc 454

<210> 70
<211> 1722
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1696)..(1696)
<223>

<220>
<221> misc_feature
<222> (1696)..(1696)
<223> a, c, g or t

<400> 70
tttggcccta ccagcccttc tcttttcttt ttcgttagct gtttgctttt tttgatccag 60

ctctgtctat attagtcctt gccatctctt ccatctgccc attaactctc tctagtgcct 120

ccgtgaggag atttcataag gacctgctag tgactggcgc gtacgagatc tccgatcagt 180

ctgggggcgc tggcggcctg cgcagccacc tcaagatcac agattctgct ggccatattc 240

tctactccaa agaggatgca accaagggga aatttgcctt taccactgaa gattatgaca 300

tgtttgaagt gtgttttgag agcaagggaa cagggcggat acctgaccaa ctcgtgatcc 360

tagacatgaa gcatggagtg gaggcgaaaa attacgaaga gattgcaaaa gttgagaagc 420

tcaaaccatt agaggtagag ctgcgacgcc tagaagacct ttcagaatct attgttaatg 480

attttgccta catgaagaag agagaagagg agatgcgtga taccaacgag tcaacaaaca 540

ctcgggtcct atacttcagc atcttttcaa tgttttgtct cattggacta gctacctggc 600

aggtcttcta cctgcgacgc ttcttcaagg ccaagaaatt gattgagtaa tgaatgaggc 660

atattctcct cccaccttgt acctcagcca gcagaacatc gctgggacgt gcctggccta 720

aggcatccta ccaacagcac catcaaggca cgttggagct ttcttgccag aactgatctc 780

ttttggtgtg ggaggacatg gggtaccacc tacacccaac aagtcaatga gggacttctt 840

tttaatttgg taggattttg actggttttg caacaatagg tctattatta gagtcaccta 900

tgacaaaaaa taggggttac ctagataatg ccaaagtcag catttgtcct gggttccctt 960

gtgtgatctg tttggactat gttttctttt cttctcccac ttgctcagca gcttgggctt 1020

ccattctagt tcttttacca agatttttgt gtgaccatgt tgacttcatt tggattgccc 1080

tctttcaatt tccttgtgaa aacaccctta actttctctt tacccttagc tgaaatgttt 1140

acatagcttc tggtgatatc ttttcatgat tttatatctc ttaaaatggt gatggatgtg 1200

acacctcata aaagtgagct ttgaactgta gataactctt aaagaaaatg tcatttaga 1260

caattaaaat atttgtgctc aactgcttga actttttttcg tgtatgtgta tttaattcta 1320

tgcaatatta tcacatgtgt agattcatgt gaccaccatc acaagagaca gaacagttct 1380

gtcacatgga tcccttgcac tgccctttta cagccgcagc cacatccctt tcttataccc 1440

tcaccccaac ctgtggctac cactgttctg tcctccatct ctgtaatttt gtcatttcaa 1500

gaatgttgta tgaatggaat catacagaat gtaatcttac gaggctgatc ttttttcatt 1560

cagcataatt cccttgaaat ccatccaagt tgttgcatgt atgaatagtt tcttcctttt 1620

tttcttttat gttttatata tttaggggggt ataagtacag gatttcttac catgcatata 1680

ttgcatcgtg gtgaantggg gggcggttcc tttttgtggc tt 1722

<210> 71
<211> 623
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (477)..(477)
<223> a, c, g or t

<400> 71
gcggccgccc gggcaggtgg gcagatcacc tgaggtcagg agtttgagac caggctggcc 60

aacatggcga aaccccatct ctaccaaaaa tacacaaaat tagtcgggcg tggtggcggg 120

tgcctgtaat cccagctact caggaggctg aggcaggaga attacttgaa ctcggagggc 180

agaggttgca gtgagccgag atcgcaccac tgcactccag tctgagtgac agagtgagac 240

actgtcttaa aaaaaaaaaa aaaagatttt tggacctgtt gttcattcat ttaagcgtga 300

attaattgtt cattttcaaa cctattttta agttattggg cttataacat ttttctgtct 360

ttcttatttt gttttttaaa agatttaccc cggaaagctt tggcgttaat ccatggtcat 420

agcttgtttc cctttggtgt gttgagacca tttttgttta tttccctggc tttccancta 480

aattttccac cacccaacct ctccgcaaga aaccaaaaaa tgggcgaaca cggcgcggaa 540

gaagaagcgc gtagacgggc gcagcggcag aggaacaaaa gcgagaacca gcaaggggaa 600

aaaaagggag agcaggcaaa ctg 623

<210> 72
<211> 1452
<212> DNA
<213> Homo sapien

<400> 72
gcgtgctcgc ggccgagtta ctgtccgctg tgccagtgcc cttgagcaat tactgcggac 60

ttcaagctca aggacggagg ccttcagtta gacaatgtag tgcccatctt taggagccgc 120

tagcgcctga acctgtgaga tgtctccacc gtcggattct cgatcatgat cccttacggg 180

gagtgcccta gattccccta cgggacccga gctcatgcat tggagggact agcatctcat 240

gaccataggg tggtcctcgc gagaaaccca gtagtctctt gcccatgtgt cttctaacta 300

gagaaccatt acagtgtcaa cctccctaag gccgttgtca agcgtacgtg gtacctcgag 360

cctcttctca acttcgttgt tgttgattag gcggtcttcc ctggagtatg ccgtggccct 420

cagtcccctc tccttaggca gataatggct tgggtatgcg cccaggtggc atttgaaccg 480

cttttgcccc taggccccga tgcgtcgtgg ctcacccccc gggcccttgg cgtgtctccc 540

gctaacgtac gccgtctttc gagcccgatg ctctcggcga cttccccgtt gtgtgcccat 600

tgcgacccca agctggttag gacttaggta ttccccacct tgcaggggac cccagggcaa 660

ccatggcgtc cacttcctgc ccaccgcttc tcgcacgttc cgacttcgct gccttctcca 720

gggtgggacc gtttccgggc acatgctctt ccaacgcgcc cccacaagca cttcggaaca 780

ctgggcgtgg tgccccattt tgaccttatt ggttccccaa cgcccacctt tggtttccct 840

tagatccaag gttacttccc ccccccccta agttggtcgg ttagaggacg cggcgggcta 900

atttgcgcgc gcgcaccccg atttctccta gcttttcccc cctttgcgtg ctctttctca 960

tttcccccat tttaccgcac gggggacaac ttatccttac agcaccggcc tttatgttcg 1020

cggtacacac gtcccgattt gccgtccagt tacggccttt cgttctcccc ctttgttttg 1080

tcttgacaca cttctggctt ctaactcccg ggcccattca caccaaagtt ttccccccaa 1140

caagcaacat acgcgccacc cggagccaca caacaaccac cccacacgaa cccgcactcc 1200

acacacccca ccccgacccc gccccctcca cctcccacac cccacaccac cctctaaatc 1260

cacccccac ccacacccaa ccatccctcc ctaccacacc actcccacac acacctcaca 1320

ccacaacaac cccacacacc agcaccactc caaccacacc tcgtacacca acccacccca 1380

ctcacacccc acaacccacc cgaccccaca ctcaccccca caccctcaca caacacaaca 1440

accaacctcc ct 1452

<210> 73
<211> 438
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (226)..(226)
<223> a, c, g or t

<400> 73
ctagtctcga gttttttttt ttttttttt ttttttttt ttttggaagg gtttaaaaat 60

ttttttttt tggaaatttt cctggaatta ttaaaaaccc cctttgggga gggaaaaaat 120

atcaccccat ggaatattgg gaaaaaaata tgcaaacacc gttgaagaaa tctccgtgcc 180

ccttctcccc cccaggggge acgaccccgt aagtaatgaa cttgtngcgt acctctgtgg 240

ctcattagcc gtgtcccccg tgtgtgttag aaagtgggtt tacccgctcc acatactccc 300

accacaacat tagcgagcac aggcctcatc acacgctcca catctactat tacatctatc 360

aatctcactc atccaccact actctcctct tctactatcc tacccacaca tcaccactac 420

ctaatccccc atctgcga 438

<210> 74
<211> 239
<212> DNA
<213> Homo sapien

<400> 74
ggcggcgcag ggtgtccgtc caggctggcg ggttgccgaa cccccgctcg ccggccgcgt 60

gcccttccgg gcatgcgctg ggccaccggg agaacgacct ttgccttggc gccgtgctgt 120

gttgtgcgcc ttgttttgcg gtccgcttcg gcggcgcacg cgcacgcgac cagtgggctt 180

ccgtgtcccc ctgtagggtt ctgtcgcacc ccggtgtggt ggactgcgta cacatgcgg 239

<210> 75
<211> 1282
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (218)..(218)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (311)..(311)
<223> a, c, g or t

<400> 75
ggggccgggc cgggcggtgt tctcagatat acaacaagat tatcgcaggg catactggct 60

gagatgcctg cgcgatcact agttccactc gagcatgtgg cgatgttgta gtggacgaag 120

tctgcccgct gagagtctca cgggacgtgt gcttgcaggt ggttcgacac gacgaacgac 180

gtcgagaaga aggtagaccc atcgggagct cccctacnac tcgcgtgtga tgagcgtggt 240

atctctcgtt cttgtacata tagtaataca ctggaataca cagattatgc acgcactaag 300

agcctaattg ntgatgtgaa gttgtcttaa agtcaagggt gctagacgtt cttggccggg 360

taattcgagt gcggtcgact acgcttgttg ttccttgctt gtgtatacat atggttgaac 420

tcgcgggcat cgagctaatg ttctcgactc acacacacag ataacgggaa ggccaatagg 480

aacagacatt cttactcgcg ggcattacag tagaccttcg aaaacacact cattgagtgt 540

ctccacgtcg ctccagcatc acatcaaacg tttgtaattg atatcggaat attcaataat 600

gggttccctc tctttcttca cattggggaa cttaaacaca cgaggtagaa aggtcacttt 660

gaagcccagt tagtattgcc attgggtgct tcgattaact tccttgaagg gtgctccttt 720

gcctgttagc aacatactct tctgcgttgg attacacagg gcatgctggc aactatccat 780

ctaggaccta aactgtattc catatttgga ttgaccaaat tggaccgttt ccaaatccaa 840

ttttattgct gcaaggcctt agaagcaggg gactggtttc caacaacact tagttagcac 900

caatttcctt tctaccctat aagcaaacaa gacaaaaaca ctaaagccct ttggtggcgg 960

cttaacactt ccattggggc ctacgaataa agccccgata ttgcccccat tggtttggct 1020

tggcacaaca cattccggtt tcaattcccc cggggctttc caaccctaaa ctttaccccg 1080

acgcaaccga gaacaatatc cccgcgaaca agcgccaaaa gcacacagcg ccacagagcg 1140

aggaaatagg agccatttgg ggctggctca acaagacaca ctggagcgta cactcgacgc 1200

catcagcacg accgcatcag gctggacatg cgacaaagca aggatactca ccctacgcca 1260

cccgccacaa ccgctccgcg cg 1282

<210> 76
<211> 1074
<212> DNA
<213> Homo sapien

<400> 76
gttctctaga tcatgctcga gcggcgcagt gtatggatcg tggtcgcggc cgaggtgttt 60

tgtcacctct ggtttataat gatcaggaag aaaccacagg atgtgaggtc caacttcgac 120

tgctcgccca aaggttgcgt aacttgtaaa agagtctcca ttcagagcat ggttgtgctc 180

ccattcccga tgctatcgtt atcttcctaa ttagactaat gatgaagcag tgtctgtaca 240

tatgcttgca actttagtat tcggccatct ttgggttcat cgtatgggtt ggtctggacg 300

cgtgtgaatc ttgcctttcg ttagggttcg ttcgttcgca gcggactgga gcccttcttc 360

cttccccaag cacgagggtt tgctcccact cttcaggcaa gtatcttgtg ggctatgcgt 420

tcgcggttgg gtcttccaca gtccttctgg gtgaccatac gtcttttcgt ccactaagtg 480

tcccagcttg ctcgctgatg ctttgaagcg catagattcc cgcatttttg agagctcgtg 540

ttcgaagcat gcactgcgct cacgttttgt gatgggtgct ttgcctcggt gagactcaat 600

tgtgtctaac atctgtttgg ttttgtttcc cctgtgtggc agtccacaac ggtacttacc 660

atgctttcct cggcacgatg gttgggtttt cgtcagttct tgtgactaac gcaactaggc 720

gcctcccttc ttgtccttcg cctcttcagg gacgctagtt ccgcgtggtc tcctcccgcg 780

gcagtgcatg gattctttcc tggaccgagg gtcggatgtg tactcccgtt tggttaggta 840

actgatatcc ggcgttgcgt gtcgatcgtg ctctcgctct tggcatccgg tcgcccgttg 900

ggagcgtggc cctacgtttg ccccgtgcct cggggtttcc gccaggcgga ttgcctgccc 960

tggcctttgg gcttgactct tcaacacaat tgcccgcctt ttttttttcc acacggcttg 1020

actcttgcct cctatctggg cctcggtagc cccccgtttg gtttttcata cccg 1074

<210> 77
<211> 1343
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (452)..(452)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (607)..(607)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (855)..(855)
<223> a, c, g or t

<400> 77
gcgtggtcgc ggccgaggta caccctgcac ccgcatcccg ggaggagatg accaaaaacc 60

agtgtcagcc tgacactggc ctggttctta ggacttcgta cctccagccg tatcattcag 120

cccgtggcag ttggcgagaa tgcaatgacg aggtcagcgc tggtataaag cagaactagc 180

gaagatccta cgactcatcc tcattgcctg gaacgtccga acgggcctca cttcttcgct 240

ctactaccga ctagaggact tcgagcctgt tggaggatcg agacgcagtc aatgtgcgtc 300

gtgccaatgg cgaagcgtgg acaacagtct tttgcgtcca atagcttccc ctggatgtac 360

ttgtcagtgt aggtgctccc ttggctagct aagccgtctg ttcaactgcc cagagacgat 420

gccgtcctcc cgtgatctcc aggggttaac anttgaagtt gctggatcgg tccggcgata 480

agtcgccccc tgtcttctcg ccctgggggt gctcgtccgg gggctcggac ggcctgaacg 540

gcaattgcct tggggcctct gataccccc ggtcctacca ttactttccc catgtggcac 600

ccaaggncca agtggggaga cgtttaacac gggtcacccc gcatgcccac tctgtaccct 660

cttaaagata tacaatataa acataccaaa atagagagga taacaagagc ccttgtgggg 720

aggagtatac cctaacgtgt ggggaagctc acccatgagg cagttgcgta tacccgcgtg 780

ggtgttggcg ccacacggtg tgctggtgtt ctattctgcc ccggttctct tggcgcaaaa 840

aatatactcc tcacntcgag gcaagaccct ttttgcggcg cgcgcaaagc tcagcacaaa 900

accccgtttt tgggaacaag ggggaacttg gggtgcaact tggaataggt ctcaggtgaa 960

ccgcacacgc ggaaactttg tagagggaaa accctgtgtg aacccacaca aagggttgga 1020

cgcgcccccc tttgtgaagt tttaagctta acccttttgg aaccccattc cgttgggttg 1080

gaagccccat aaacccgttg gagcttaggg ggaaactttc cccgagggaa gcagagtttt 1140

tgtataaaac ccaacacaca acaaacgaca aaaaaacagc aagacaaaaa cacaacagaa 1200

gaaaacgaga aagcacagac agttgtaggg cagagagaaa cccaccgcag tgggactcac 1260

acacagagcg tgtgtgtcct ccccagcagg gtgtgtagta aaaaaggcgg ggaagcacgc 1320

cagcggccac agcacagccc gcg 1343

<210> 78
<211> 1530
<212> DNA
<213> Homo sapien

<400> 78
tttttttttt tttttttttt tttttttttt ttttttaatg gggaaaaaat tttttttctc 60

tttttttaat tctcttggcc aaaaaaaatc ttcctccaaa aattattttc ccccccacaa 120

agagtttaaa aaaataaaga aaagcagtct attgggtcgg gccaaatagt ttttgtggag 180

acacctcctg ctgtgtttaa cagaggagag agagctcctc ttgtggcggg cggtgtggtc 240

cccacaccag agagagttct cgccgcaaga cagagaggag aagagatatc tctccctgct 300

gagacccgta gatatatatg tcgcacgaca cacctctctt gtgtggagtg tgtcggcgcc 360

tctccgagag aaaaaaaaaa acgcccaaat atctctctgg aggcgggaga gacacgccct 420

attatctccc actacagagc actgtgtggg cgcacacaga gtgtgtgttc tctctacaag 480

taagagacat ctctctctat aacacatatt cacactctac taagaggaga tatatatctc 540

tgtacagtat gatagagatc tctgtatacc ataattatat ggtgaggcag ccataatgtg 600

tgtagatgta tatcccacag tgtgttttaa caaaagagag accaactatt tttttctctc 660

tcctctctga agcggaatat atattgtgtg tgtagtagag cattatactc atctataagc 720

agaccacaca gctcgatgtg agaagaaaat aataacacct ctcatcttga agtgtttctt 780

cacaacacaa acacactgtg tgggggggga gaacactctc tctttatgtg tgtcgtcccc 840

cccaaggtgt gatctcccgt gagaaagtgg gggcctcccc cccccaagag aaaaaggcac 900

atatctccct ctccctatca cgcgtgtgtg ctcggggaat tctcaccaat atatctcttg 960

tataaaacac aaagatgtgg ggaccaaaga ggggaggaga cacacaaaag attatgtgtg 1020

gaggactatg tgaacaacaa taaagagcgg cggggggggg gagataaaca caccatagtg 1080

cgccgccaat agagagtgag ttaaccccgc gggtggtgtg agaacacatt gtgttatatc 1140

acgccggaca cacaaaaatt cccccccaaa tagatataca gacagtaaat gaaaccacaa 1200

aacagtgaga tacaccatga caaagtaaca acagaccaca ataataacat agagcatgtt 1260

gaaagaaaca cacaactacc agctgtaaca aaacaataaa caaacagaga gacacaccac 1320

aaaaaaacaa ccagcacagc acagagagta gcacaaagcc gaagaccaga gtactcaaga 1380

caacacacac aagaaacagc acggagcaga ccagacagga ctgacgagca agcagcgaca 1440

aagtgagaca agaacagcag caacaatcac acgaacacga acaagaagcg taatgaagca 1500

gatcagctga aaggcaagac gcacgagaag 1530

<210> 79
<211> 1428
<212> DNA
<213> Homo sapien

<400> 79
tttttttttt tttttttttt tttggaaaaa cttttaataa tggtaatggt ggttgggggt 60

acaggggtga tgtccaaatg cccaggaggc cataggggtt atagggcaaa gggggacgca 120

caaaatgttg gaaagataga gggcggctgg ccatcaaagc tgggggggctt caggccaaaa 180

acagggagct ggagggaacg ccacaagggg agggacattt cctggcgagt tggcgtggaa 240

cccactgtaa gcaacccagg tgtccctgga gaaaagcgcc ctccttctgt ggtgggaaag 300

atataacaca cgcacacgga gggagaaagc gtgtgggggg tgatattcac agtgggtcgc 360

atacgctgtg ttccctgggt gtgtgagaat atgtgtggta tatctcgcgg gctctcacca 420

atgtctccca ccaccaacca ttagcgggac gaaccaaagg agaaaaaaaa aaaaaaaaaa 480

aaaaaaacaa agaagagagg aaaaaaagaa aaagaaaaaa aaaaaaaaaa aaaaaaaaca 540

aaaaaaacaa caagaaaaaa aaacaaaaaa cacagaacaa caacaaaagc aaaaaaagaa 600

aaaagagaaa aagaaaaaaa aaacaaacga aaaacaaaaa acaaaaaaaa aaaaaacacc 660

aacaagcaaa aaggaaaaaa aacacaaaaa caacaagcga aaacccaaaa acacgcaaac 720

aacaccaaca caacaaacac caaagaaaac aaaaaagaaa aaaacaacaa aaaaccaaaa 780

agacacaaag agacagaaca gaaaaagaaa aacaacaacc ccaaacagaa caaaacgacg 840

accaaaaaca tagaaaataa aacaaaaaaa aaacacaaag aaaaaaaaac aaacaaaaga 900

aaagagaaaa gaacacaaca acaaaagaaa acgcaaacaa aacaaaacag aaaaaacaca 960

aacccacaaa aacaacacaa aaaaaaaagc aacgacaaca caaaaaaaag aacacaaaca 1020

aacaacacaa aacaaaaaac aacacacaaa acaacacacg acaaaacccа acacaaaccg 1080

aaagagagca aacaaaagca aacggacaaa cacaaacaca aacaacacaa gagcaaaaca 1140

aacataaaaa agcaaaagca aacaagggag acacacaaag caaacccaac acacaaaaag 1200

aaaaaaagac cagacgaacc aaaccacaga aacaaacaga cagaacagaa ccaaaccagc 1260

caaaaacggc agaagggaaa caaaaacaga gacagacaca acaaaaaacg aggaaagaca 1320

aaaaacaaca aagcagaaca aaaaaaaaaa cagaaaaaca agaaaaatga gagagaacaa 1380

gacaaaaaca ggaaacgaca aaagcacaga taggacagaa aagaagga 1428

<210> 80
<211> 1581
<212> DNA

<213> Homo sapien

<220>
<221> misc_feature
<222> (351)..(351)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (692)..(692)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (778)..(778)
<223> a, c, g or t

<400> 80
gcggccgccc gggcaggtac attccgtatc tatttttttat tattatttat atatatttat 60

ttggaaacag tctcgctttg tcccccagtg tgggattgca gggggccgca atcctgcggt 120

ctacagctgg gaacctcacc acctcccgcg ggtacacagc gcccatattc caccgggcct 180

acgagcccac cacgaagtaa gctcggggaa tacagcgtgg gcgccgcgac caccagcgcc 240

cgcggacaaa tatatagggt aattataaga gaaagacaac acagggggtt ttcacacccg 300

gtgtgtataa accacaggag agtgtggctc tctcagatat cctcctgtga ncctgtgtgt 360

gtgatctcac accccacacc ctctcgggca ctctcacaca aaaggtgtgc ggggagatat 420

cacacagggc ggtgtgaccc gccattgtgt gcgcgccgcg ggcaccaata tatccagtgt 480

tctcttatat cacaagagag aaatataaaa accccacaga gatatataaa tatctctgtg 540

agaatcgtga gatcaccaga tatatatagc ggccccaata tatatataag agtgtagaaa 600

accaatatct ataaaagagg aaatgatgga gctatctcta taaaaatatt aagagcatct 660

ctattgagcg cgacgaaaat ataaccccac gncgtctgtg tagtgagaac aggggtaata 720

tctgaggagt gcgctctaaa caatatcgtg tgtgcacaaa gtaaaaatat ctccccanaa 780

aaacgcgtgt aaaaaatata acacccccct ctttttgtg ggttgggggg gccaaatatt 840

tgagaaaact ttttctacac acaggggga gtattcgaga caattcccca caatgtgatt 900

agagcaacca cggaagtgtg gccctagaat attgggagag aaactttggg gagtatctcc 960

ttaaaaaggg tgggttgtat ccatttttcta aaaaatgggg gcgtggtccc cttaaaaatt 1020

tggggataaa catttaaaat accagggtta taagtgaatt acacatccgg gaggggagc 1080

aaaagggaag cctaacaaga agtttttttt gaaacaaata ataatccaaa atatataatt 1140

tcccaatgtg tgttgcaatg tattttgcta tttgatatgt gataaaaaaa ttaaaaaact 1200

tttctaaatt aaaggggggt ttggtgcaca ctggaaaaaa aaaaaacaaa aaacaaaaaa 1260

aaaaaagacc tgtgtggggt caacacaggt ggcaacaaga agatgacccg cctgggaaag 1320

cagtgtgtac cgcccaactc ataaaaaatc aaaaagccaa caaaaaagaa agaaagaaaa 1380

aagaacgcaa gacaagcaaa gatagagaag caagaataga gataaacagt aagcagacga 1440

atgccaagcg aagagataat atgagatgca gcaagacaac agagaaaaga aacacacgaa 1500

cacaacaggc ggcacacagg agagacgcga caacaaacac tgcagcaaca actacgacca 1560

gcgagagcaa aaacagacaa c 1581

<210> 81
<211> 769
<212> DNA
<213> Homo sapien

<400> 81
aatgccatgt cgagcggcgc agttgtgatg gatggtgggc cggcgccgac ggtacttgtg 60

cagtccagat atatgtctgt cggtttggat gttgttgcag ttcagatcct cgtcagcaac 120

tcctagcgat gctgcaccag aatggactat gtgtgcagta gatccatgtt atgtccatag 180

gattccatag acgaactttc ttcaccgatc gcgtgtgtga gcttcccata cttatggtct 240

atgcgcttgt cgcctcagaa atgaggatcg tattccttca tttcgttagg agctcgacga 300

gtcatgaggt tattggtgga catcagtgga ccacagggct gtgagtgata atcttgagtc 360

gattatcata gtctacgcgt tgtcgtgtag gactcgatgc tctcagtctc tacccattca 420

ctcttatcag aatacccgag gcatggcgga ccacttaagg tcaggagttc aagaccagcc 480

tggccaacat ggtgaaaccc cgtttaatat tttacattaa aatataaaaa ttagctgggc 540

atagtggcac acgcctgtaa tcccagctac tctggaggct gaggcaggag aattgcttga 600

acttgtggag gcagaagtta cagtgagccg agatcgcacc actacactcc aacctaggca 660

acagagcgag actccgtctc gaaaaaaaaa agatatcaaa aaaaaaaaag gttgggggta 720

acctgggcca tagtgtccct gtgtgaattg ttttccgccc catttccca 769

<210> 82
<211> 679
<212> DNA
<213> Homo sapien

<400> 82
gcgtggtcgc ggccgaggta ctttggcctc tctggagata gaaggcttat tcagcagagc 60

acacaagcag aggaaggtgg ataacgccct ccaagtcgag gtaactcccg agcgacgagt 120

agtcaccacg taggacagcg acacgcaaag gacaagctac cgtacgaaga ccatcaagac 180

gatgctaccg cctgtagcgc atgatgcaga acgctacgga acgtacgagg aaatcagcaa 240

caggtcatac agaccatgct gacagtcagc cgcaatcagt ggccatgaag gcgtcgacac 300

gcggtcacca gaaagatgcc ttccaacaag gaggcgcacg acgtgatata acgagcgaga 360

cgcaaatgat cgtacaccgg cacacgtgag atcccctcta ggatatcgca cgaacgtgga 420

caccctcatc acccatatct catgttgcga ccatcgctga acctcatata tatatcgcca 480

gcgacgtgga gaacaccata aatccccctt acttcagcgg ggtgcccttc gcaaaagtct 540

tacaagtcta ttatcaacac gtacaagcgc accacacttc acttcacatc tcatataggg 600

cgtaataaca tttagttgct aaaatgtatc gagaagggaa gacatgcaat taagagtaaa 660

agtgcaattc ttttaagaa 679

<210> 83
<211> 1180
<212> DNA
<213> Homo sapien

<400> 83
gcggccgccc gggcaggttt ttttttttt ttttttttt ggaaagagaa accccggtat 60

tgattgtcgg ggttagagga gatagagagg aaaatgtggg ggaataggtg tgttttagac 120

catgtgaggg tgtgttttcc ctcggtggtg aagtgagggt ttaagtgttg tgtaagtggt 180

gcggtgcggt gtgagtgtag acgccccatt gtggtgtgtg tggtggttaa attatgtgct 240

aagagaggca gtatattgag ggcgtgtgtg acacatagat gtgtgtgtgg aagtctccgt 300

gtgtaagttg aggaagaagt gtgaatatgt gtgacactcg aggaaaacac accggtgggt 360

ttcacctaac ccaccgagaa gagagttctc ctcccagagg tgagggttta atatagggtg 420

agaggggat atagagcgcg cgaaagagta taatagacag agaaagaggc cttgttgctc 480

ctataaaaag aagctctcta tctacaaaag gaggggctat atatagtagg ggggagaaga 540

tatagaagat ttgtcgaaga ggtctcctgt gtagaaagag agggaatttt cagtgagagt 600

ggcgcacagc gtgtgttgaa agtgtgcgcg tatacccgac aaagtaatgt gagagagtat 660

gtgtgcgctt gagcgcccca agaaaacaca tagtctcaag tgacaggaga tagttgttta 720

ccagcccccc gtgggggggc cgcgacaact tctagtggcc accacactgt tgaaagctgt 780

gtgccccccg cgtggttgta aaaagcctgt gggggggagcg taataaaccc tccaagggtg 840

ggcctccaat ataagcgcgt tggtgatacc ccgccgtggg ttggtgttgg cacaacaagg 900

tggtggcgtt attaaccatc cgcgggctct tctacaacca agatggtgcg tcccccagac 960

gagacaaaac atttggtggg ggggtggagg ctacaccaaa caaagtggag agtggggaga 1020

gagacacaga cagtaggcag atagggagag aagaggagta caaagacgag agagcaagag 1080

aaagatagaa gaagagagaa gaagagaggg gaagagaagg cagaaggcag acagcgcagg 1140

aagagaggcg aacaggcaag aaagacagca aagagacaga 1180

<210> 84
<211> 516
<212> DNA
<213> Homo sapien

<400> 84
ctacggactc gcgcgcgacg cggccactca gctcactcaa caggccggcc cagcacgcgc 60

catcccacag acaccagggg atcaacgcca ggacaagacc catgtgagca caaagggcca 120

cgcaaaggcc aggacccgca cacaaacggc cgccgctcgc tgcccggtta accactaggg 180

ctccgtcccc ccctgtacgg ctgcagccac aggtacactc gatcgcctca cgttcagagt 240

ggtgatcgca tccgacaggg actataagag cttacccagg cggttgcgcc acatggaatc 300

caccctacgg tgcgccactc ctggtcccga caccttacag cacaccggag tacctggccc 360

catcacccat cgggaacagg tggggtccta taccacaccg ttacggatac cccccgcggc 420

agccgacagc caaacggctg tgtacaaccc actccgccgg cggcgcccac acagagcaac 480

gcctaggaaa ccaaaaacaa ttacacgaaa aatgaa 516

<210> 85
<211> 669
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (421)..(421)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (538)..(538)
<223> a, c, g or t

<400> 85
gctttttttt ttttttttt ttttttttt tcccaactcg gggggcaata tttttaatta 60

aaaactactt tttatattta caaagatgct attgaacaac aaaattataa tggaccttaa 120

aaaacccccg agggaaacag aaggtttcaa tttttcagaa atcctaaggg gggccccggg 180

cgggggggcg ggggctgggt ggcaccggca gtggttcccc ctggtggtgc agcaccttgc 240

agactggccg ggacttcgga gaccaaaagt gactccacaa tagcaaagac cctccagctc 300

tttcctgtgg tataattcca ttctccccaa cccccagtgt gtccgtgtgt ttccctgtgc 360

cgctccccac gattcgcatt tctccaggga caacggtccc tcacgttaac tcggtcgggg 420

ngagtacact cgtggtcctg gaaagagtgt gggggataca ctcgtgtggt gctcaatagc 480

cgtgagttct cgcgtggggg gtgaatagat gtgggttact cgcgctctcc aacaaatntc 540

ctccagcaac aaaccattac cccggaagca caaacggggg gcggcaaggg gagcaacaac 600

caagccaaca caacacaaag aggagaacaa ccaacaatat gcaacaacaa agcaacaaga 660

aaacaagaa 669

<210> 86
<211> 371
<212> DNA
<213> Homo sapien

<400> 86
cgagcggccg cccgggcagg tgctttttt ttttttttt ttttttttt ggtttatgct 60

taaatctttt tttttcagcc tcaggggtg ggggtggcgt ggagaccatg tgaacttctc 120

aggtctccag agaaaatgtg ggtttgtgga tctccagctc tttgtgccat ttgtgctctc 180

tctctgtgaa ataaacacct caaaaacatt tacacctcct ctcttaagcc gtggggcgta 240

tatctctcgt gtggctcaca atagccgtgt ttccgcgggt tgtgaaagtg tgtttacccg 300

cctcccaatt ccccccaaaa catccaagaa aggggtaccc acaaaaggaa caaaggagaa 360

aaagaaacca c 371

<210> 87
<211> 998
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (332)..(415)
<223> a, c, g or t

<400> 87
ttagatgcat gctcgagcgc gcgcattgtg atggattggt cgcggcgagg ttccgacagt 60

cagccgcatc ttctttatgc gtcgccagcc ggaggccaca tcgctcagta caccatggta 120

gaaggtgaag gtcggagtca acaggattta ggtcgtattg ggcgcctggt tcaccaggac 180

tgctttataa ctctggtaaa gtggaatatt gtttcgccat caagtgaccc ctttcattga 240

cctcaactac agtggctttg acatgttcca atatgattcc gatcccatgg gcaaatttcc 300

atggcgaccg tccagggctg taagaacggg gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnaggta 420

tcatgagtat atagatcgtt aatccccacc agactagact ttgaacttta gtcagacttg 480

aagatttggg cacgttatcc ggtgtcattt actaggacgg tgcatttctt gcgttctgtg 540

atgactgtga tcttctcctc agcacgaaga cgttgtctgt attgtcggca gggatacggc 600

ctcactcacg actttccttc ggtcttcctt tgttctcccc tttaagctcc gtttgatttc 660

aagctggtgg ttctacgggc atcttggggt ttctcccccc attcagtgtt attctcggaa 720

tctgcgtttg tcagcttcgt tgatgtcctt ttaggcctaa tattccaatt gtttggcctc 780

ggggaaccct taacatgttc ctaatgactt tagtgtccga caagcttggc cgtactcccc 840

gtccgttgcc tgttcctgtg ttggccttgt ttccccggtc gcgatccccg catttccacc 900

aatccggggt tccccctgga gcccccggcc cccgttccac ctccgcgtac cggacgcgcc 960

ttcccccctgg cgtggcttta ctctcccctc ccggtccg 998

<210> 88
<211> 457
<212> DNA
<213> Homo sapien

<400> 88
gcgtggtcgc ggccgaggaa cttatacccc ctaaatatat aaaacatttt taaaagaaaa 60

aaaggaagaa actattcata catgcaacaa cttggatgga tttcaaggga attatgctga 120

atgaaaaaag atcagcctcg taagattaca ttctgtatga ttccattcat acaacattct 180

tgaaatgaca aaattacaga gatggaggac agaacagtgg tagccacagg ttggggtgag 240

ggtataagaa agggatgtgg ctgcggttgt aaaagggcag tgcaagggat ccatgtgaca 300

gaactgttct gtctcttgtg atggtggtca catgaatcta cacatgtgat aatattgcat 360

agaattaaat acacatacac gaaaaaagtt caagcagttg agcacaaata ttttaattgt 420

ctaaaatgac attttcttta agagttatct acagttc 457

<210> 89
<211> 3100
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (49)..(49)

<223> a, c, g or t

<220>
<221> misc_feature
<222> (91)..(91)
<223> a, c, g or t

<400> 89
gtctggtttg tttggcccac cagcccggcg cggccctttt ccgttagcnt tgctgctttt 60

tttcctgctc ggccccagat tggtccttgc natctccttc catctgccca ttaactctcg 120

caagtgcctc cgtgaggaga ttcacaagga cctgctagtg actggcgcgt acgagatctc 180

cgaccagtct gggggcgctg gcggcctgcg cagccacctc aagatcacag attctgctgg 240

ccatattctc tactccaaag aggatgcaac caaggggaaa tttgccttta ccactgaaga 300

ttatgacatg tttgaagtgt gttttgagag caagggaaca gggcggatac ctgaccaact 360

cgtgatccta gacatgaagc atggagtgga ggcgaaaaat tacgaagaga ttgcaaaagt 420

tgagaagctc aaaccattag aggtagagct gcgacgccta gaagaccttt cagaatctat 480

tgttaatgat tttgcctaca tgaagaagag agaagaggag atgcgtgata ccaacgagtc 540

aacaaacact cgggtcctat acttcagcat cttttcaatg ttttgtctca ttggactagc 600

tacctggcag gtcttctacc tgcgacgctt cttcaaggcc aagaaattga ttgagtaatg 660

aatgaggcat attctcctcc caccttgtac ctcagccagc agaacatcgc tgggacgtgc 720

ctggcctaag gcatcctacc aacagcacca tcaaggcacg ttggagcttt cttgccagaa 780

ctgatctctt ttggtgtggg aggacatggg gtaccaccta cacccaacaa gtcaatgagg 840

gacttctttt taatttggta ggattttgac tggttttgca acaataggtc tattattaga 900

gtcacctatg acaaaaaata ggggttacct agataatgcc aaagtcagca tttgtcctgg 960

gttcccttgt gtgatctgtt tggactatgt tttcttttct tctcccactt gctcagcagc 1020

ttgggcttcc attctagttc ttttaccaag atttttgtgt gaccatgttg acttcatttg 1080

gattgccctc tttcaatttc cttgtgaaaa cacccttaac tttctcttta cccttagctg 1140

aaatgtttac atagcttctg gtgatatctt ttcatgattt tatatctctt aaaatggtga 1200

tggatgtgac acctcataaa agtgagcttt gaactgtaga taactcttaa agaaaatgtc 1260

attttagaca attaaaatat ttgtgctcaa ctgcttgaac ttttttcgtg tatgtgtatt 1320

taattctatg caatattatc acatgtgtag attcatgtga ccaccatcac aagagacaga 1380

acagttctgt cacatggatc ccttgcactg ccctttttaca gccgcagcca catccctttc 1440

ttataccctc accccaacct gtggctacca ctgttctgct cctccatctc tgtaattttg 1500

tcatttcaag aatgttgtat gaatggaatc atacagaatg taatcttacg aggctgatct 1560

tttttcattc agcataattc ccttgaaatc catccaagtt gttgcatgta tgaatagttt 1620

cttccttttt ttcttttaaa aatgttttat atatttaggg ggtataagta cagatttctt 1680

acatgcatat attgcatcgt ggtgaagtgg gggcagttcc ttttgattgc tgagtagtat 1740

tccatggtat ggatgtacca cagtttgctt aaccattcac ccactaaagg acataagagt 1800

tgttttcagt tttttgccct aataaagctg ctgtgaacat tcatgtacag gtttttatgt 1860

gaacatacat tttcattttc tgggataaat gctcaaaagg gcaactgttg ggttgtatgg 1920

taaacacata tatttttgta agaaactacc ctactctttt tccagagtgg ctctacttttt 1980

tacatacagc cactcataca attcagacag caatgtatga ttgatccagt ttcttcacat 2040

cctcaccagc atttggtatt actactattt tttatcttaa ccattcacat agatgtgtgt 2100

aatgatacca catgtggttt taatttgcat ttccaatggc taatgatgtt gagtatcttt 2160

ttgtgtgcta atttgccatc tatgtatcct cttcggtgaa atgtcttcat gtcttttgtc 2220

tattttctat ttaggtcatt tgttcttttt actattgagt tttgagaggt tttttatata 2280

tcctagataa aattcctctg ttagatatgt ggttgcttga atttttaaca taacttctac 2340

caaggaaaaa taagtaaaat ttccaaccct tgcatggcca gtcacttact taattcctgt 2400

ccttcagtgt tccatctaga gaattaagag atatgatgta taaaatagac atcgagggcc 2460

attaagagag taaatactta aaaatacatg ttatgaaagc aaagccaata atcactgtag 2520

gagtatgagt tgcctaaggg ccaaaactaa tgtaaataag agaaagtgtg gatataaatg 2580

accattgttt ataaacagtc atgaaaaatg ctgtgacttg aaatctttcc cacatctccc 2640

aagaaagtag gtaggagttt atcctttccg taatctcttt ttaaccctgc tgactattac 2700

agggcttgtt taatcacagt ggcaagaatt acatgtatct tacagtaaag aaacagaata 2760

ctggaatcgt tagagaaccc tgatgtgttg acctggataa agtacaaagg tggaagaggg 2820

aatgagttat gctgttaaaa tctcaggcta ttctgttaat gttcctgcta ctatgaaccc 2880

aaacttttt tttccccctt ttgactcctt gtgtcttcct ctcctgaggc ataaaagtag 2940

ttctgtcggg ttaactgtca caccattggc acctgcggcg gcaactggga ggacatgggc 3000

tgaaaaccgg gcggggagcc tagaaagcgc cgaatgtgga agccttctcg ggctacttac 3060

acacggagga aggacctgga caaattggga aacataatag 3100

<210> 90
<211> 1304
<212> DNA

<213> Homo sapien

<400> 90
tactcgaact gaccgttgac tatgacactt ttcggctcct gtgaacgttg ctggccccaa 60

tgctgtactg atccatctag acaccagagt agactgttca taggatgcct agatagggta 120

ccacgagacg aatcgttccg gctgacattg ttggattacc cctggggtcc agtagcctca 180

ttcgctctag cgagtcgact accgttgcgt actgggtcat cttagtacta gcgcgcatag 240

gatcatggaa cagctactgg cgttagatca atggagtcag tgagccatgc tctctcgctg 300

gtgctggaca gttgtagcta tatgccgact tccagccaca ctatccgact accaagcact 360

aacctgagac aagagccaca gctccctcac ctaggactga tatcagatat ggggaacgta 420

tgacactttc gcaatcacac atttgagcga cagggacggt cactcatggg gagcacttac 480

gtatgaatga agcccaacag agagacggta ggagggatat aggaggcgcc gctgccatat 540

gtctggatta atccagtttt agtggctcaa ccaaaggacc tacatatcag cgcatatggg 600

ctacgttggg aatgaaggcc acttcatata cccttcagag atccgcgcta tccctagtgt 660

gggatgccca acacattgta ctacacaaca cgcataggca ccatagcgca aagatcccca 720

catagaccat tcgtcacacg agtccgggct accgtacccc atagactcct acctccattc 780

gcgatcctac cacctggctt gtacccgtac cccctgtgtg gtaaggcaac agaggataaa 840

ttatcctaaa aacttggatc ttaggtgccg cgttgttggg ggcccaccaa ggaatccacg 900

aaggggacat gaatcgacaa accgatatta gaagagctac ttttatccct tttaagaagg 960

ccctttcgac ggtcccctcc ttttcggata taacccgggg acacatccga aaggataggt 1020

gcattataaa atcttcccaa agagccccgc ggacaatcgc acatgcggaa ccatatcgac 1080

cgatatgcat attacccccg tggtgctccc cgggagagga gtcgttggaa caagaggaac 1140

gggggaacaa gaaagaaaca acttgcatgc gcctgtatac cctaaatcct ttgacaagct 1200

ttagcactgg agacccctcc cggaaggcat gtagattcgt gagagacaaa ttccataaaa 1260

aagcgtccct taccaatgtt tgttccgatg ggactataga gggg 1304

<210> 91
<211> 993
<212> DNA
<213> Homo sapien

<400> 91
gtggtcgcgg cgaggtgctt ttttttttt ttttttttt tttttggttt aaaggtgaaa 60

aagccaaaaa ttttttttt caattgaagg gaaccataat cccccgtgtt gggttacccc 120

agggaattcc acagtccatc aaaaggaacc attccaaata atagctaaaa atctgatagg 180

ccgcggacca gtggggtcta ttatgggcaa gggtgtgctt tcacccccaa atgtttcctg 240

gagaaaaagg atgcccagtg tgctccgggt ggcgcaaaaa gaacaagggc aggaaatgtg 300

gtggggtggg ggagaggcgc ttgggttgag aaaaacacac tggagacgca ggacgcaggg 360

tgtcactttc tgatctccca ggctctggaa tatgcgccat gtgcgcctgg gcacatatat 420

aagggacaca aaaatatcgc gcttttgtga acatatcggg agatgttgtg gggctggggg 480

ccgcgtgtgg cgctccagcg ccttgtatat tccccagcat ctttgggggg cgccagggtg 540

ggtgtggtga tccacggatg gtccaacgag tatttgacga cctatccggc ggtcttaacc 600

ccgtttgaac cccgcgtcct ctatctaaaa aaatattccc caaaacaaac acaaattttt 660

gcggcggtgg gtgggggggc gccgtttggg ttgtccccca gatatcccgg ggggtgtggg 720

gggacagaaa agtggggggt atgcccgtgg ggcggggcgg ccctttttggt gaagcccgaa 780

aagtcggcgt ctctttgggc cccacgcgtt gatgtataag ggcgaaggcc ctctctttac 840

aaacaaacgg gcgcgtgctt ggctctcagg ccccaggggc gaggggaggc gtccctacag 900

ctgtgcgcgt ggacaccctt tggcgtccct gttgtgtaga gctctggggt gttgtgaggc 960

gcaattggct gctgtcttac cacaactctt gtt 993

<210> 92
<211> 1439
<212> DNA
<213> Homo sapien

<400> 92
tggtcgcggc cgaggtgctt ttttttttttt ttttttttttt tttgtggttt aaaggttgac 60

accacccaaa atctttttttt tttacaaccg tgacggggcc cattaagtcc cccttgttgg 120

ggtatacccc agaggaaagt tacacagttc tcactcaaaa ggacaccaat tctcaaatta 180

atatagcata aaaaactcgt gaatgaggcc gcggaaccag tgtgggccat atatatgtgg 240

cacacgagtg tgggcacatc tctacatctc tcccacacag agttctctcg ctgtggaaag 300

agaggtatgt ctcccacacg gtgtgccctg tgggtggcac acaagaacac acagaggcgc 360

gaggaaagag tgtggtgtgg tgtggtggat gaggggcctg tgggcgagag agagagacac 420

acaccaccct gtagagagag agccggagag acacacaggt ggtcgtcaca cacttgtgag 480

actctctctc ccacgggggg ccgagaaaaa tgtgtggcac cagggtcgcc gtgcgcacac 540

actataaaag agggagacga cacacaaaaa acaatgtgtg cgcgtgtgag aaagacaata 600

gtgagagaga gaggtgtagt gggcgcgagt gggcgcgcgc gaggggggcg cctcacaaga 660

gcgccggtga gaaatctccc agagaccact tgtgtgtgga gaggcgccca caggggcgg 720

cggggagact ctcacagaga gtgggtctac aaagagagat gtgtgagaca cacacaactc 780

gcgtgtgcgc gtaacacaca cgcgcggaga aaatacccgc gcgggtctct ctctacacaa 840

taaaacaaca tatctcaaaa aaaaaaacac aaaacaaaca catagtgggg ggcgggggga 900

gggggcgggt ggcgcgccac actagttgag agacaaatct cccccagaga aaacaccgca 960

ggggagagag cgcgatgagg gcgcaacaaa caaaatactg tgcgcgagta accaccccgg 1020

ggggggagag acgcggagac accttctgaa aaaggagcgc acccgaaaat aaaccccggc 1080

gctcatctcg agtggtggca gcccaacagc caacgcgggc ggtgtgaagt taaaccacca 1140

gtagaccgca cagagaacaa caccccacca acactcccac cacaaaaaac aaacaaaaac 1200

acacacaagc acaaacacaa acgagatgag ggaacacccc acaccaaaga aggaagctca 1260

cccaccagaa gacaacacag gcgcggacga gaagcgaagc agagaaccac atatcaaaca 1320

acatgcgttg gtgcaccaac agcgggggtg gcaaaaccca aaccctcgtg tggtgagaca 1380

gaccaccacc cccagaaggt gggcgaccaa aaagaaaacc cacctcgaag gagaaaatg 1439

<210> 93
<211> 889
<212> DNA
<213> Homo sapien

<400> 93
ccgccgcggg caggtattct gctacaaacc agcaaagaca ttggaacact atgacccgta 60

ttatgtcggc gcatgagctg aactcctagg ctacagctct atagcctcct tattcgaggc 120

cgagcgtggg tccagtccga ggcgagacct tgtctagggt agagctacca acatctacaa 180

cgtctatcgt cacgagccca tgcgatttgt aagtaaagtc ttcgttgcca tagtgaatta 240

ccgcgatcag taatgcgtga ggagcatgtg gcgtcaaccg ttggacctag gttctccgcg 300

taattaactc gggaattcca ccgcgactac tgtgcggtta ttcctctctg cagtaaatga 360

caatcagtat aggcctctgc gacgtgtaca atcagtggac gtcattacgc cttgggtttc 420

cggtgattgg ctgtaaacag tatgcatgta gctccggctt cacagacatg tatccatgca 480

gcacatacat tagcggaagg ccagcaaata aaccgagtgg gaatggatgg cgacgccgag 540

tggcgtatgg tagacgccgc ccaggggact cgagcaggga aaacgaaccg gccattacca 600

ccgtcggaat agtacactca aagcggaaca agcctaggtg gcgagagctc aggatcccag 660

catgaaggac cagtgaacgt cccctggagg acactccgtc ggatagctcc gaaagggcct 720

ttgtgtggac gtgctacaag ttactaccta tgcgggatcc caatgctaca gccaattgga 780

tacctcccta tgtggtaaga gttcgaaaga caccattggg tataagatca cctggctgtc 840

gaaagcgaca gtttccccaa ccaccagaat cactatccgg aatggccaa 889

<210> 94
<211> 626
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (176)..(176)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (246)..(246)
<223> a, c, g or t

<400> 94
tggcgaatgg gcctctagat gctgctcgtg cggcgcctgt gtgatggatt ttttttttt 60

ttttttttt ttttttttt ttttttttt ttttttttt ttttttttt ttttttttt 120

ttttttttc ccccccgggg tgggggggg tgggggaaaa taacctttgg gcccanagaa 180

aagatgtggt gttgtggtcc tcccgctagt aggaccatat gtgtgctctc ctcgtgaaaa 240

taaccncctc tctacacaga agatattcac tcctctctct ctctctatac aggctgtgtg 300

gagcgtatca cgccagagtg agtcacaaat gagcgagtcg tgttcgcgct gtgtgagtgg 360

gaaatagaga tgtgtgatct catctcgcgg cgcttctaca caaccatatc atctccacac 420

aacacagcac agcactacta gtacgacgac acgacgacga cgagacgacg gggaacaaga 480

gcagagacaa ggagcggcag cggagagcag agaagaggcc caacagaaaa gcaagaaaga 540

gaggaagaga ggcgcaaagg agacagaagg acaaacgaca aaagaggagg cggacgagag 600

aagaggaggg agggaagggg cgaaac 626

<210> 95
<211> 507
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (98)..(98)
<223> a, c, g or t

<400> 95
atgggccatg gcggtagtga agctcccagg tgcctatgag gcaggggttt ccacctgaat 60

gagcatctct gatagcatga tcatccgacg ctactcanga tgaaagtggg aacatggatt 120

ggacaacgtt catcagagtg gttggtcagg ctgtacggcc aatgttccgg tgccttcaaa 180

tctcttgatc aaagatggcc acgtggacaa atcctagggt gaacgaatcc agtgagtggc 240

atgcagctga cgtatgtatt ccttgaacat ggaagcctcc aggagcctga ccccgtcacc 300

gactatggta taggagatca cctggagccg tgccgcatcg tcatccgcag ggaccggatc 360

gctgttctgc gagtctacat ctttgtgggg cctgcttgat gacagaagca ttggcgataa 420

ctcgtggtca atagtctgta atccgttgtg atgaaaatat gttgatccta gttcacaatt 480

tcacggaaga atataaggaa ggcaagg 507

<210> 96
<211> 1074
<212> DNA
<213> Homo sapien

<400> 96
cgtgatagat cactataggg ccattggtta tctagatgca tgctcgagcg gcgcatttgt 60

gatggatagc ggcgccgggg aggtacatgt ggaaggtgga taacaggccc gtccaatcgg 120

gttgctgcct ccaggagagt gtcacattgg atcaggacgg ctggactagt cacctacgag 180

cctcatggca gcagcctgac gctgagcaag gcggattacg agaaacagaa taggtctacg 240

ccgtgcgaat tcacccgatc acggtgccgt gagtttctgg ctccgtctct atgagcgttc 300

aaccggggag aaggtgtccg tagggagtac gttgcccccc actcgtgtgt ccttcgagct 360

tccagcctgt tgacctcgcc gtcccactcg gtttggttcg tgcgtgcacc gcttatctcg 420

cacaggggca tctacgccct atctggcggt cgtccaggct cattcattgt ctcgcgtcca 480

tcctccttcg ctcctcgctt ggcttctgtg tttgcttgtg ttggcgggac gacatgtaac 540

taacaataaa gatgacggtc gtctatgccc aaacaaacaa aaaaaaaaca aaaaaaacaa 600

aagagactcc gtggggcgtt actccctatg gtggcccaat gagacggtgt gtctccccgt 660

gggtggttga aactgtgtgt gttctctccg gcaacaccaa tattctcccc ccgcacaaca 720

ttctccgacg accccaaacg cagaagcaca ccacacacaa cccacgaacc acactacaac 780

aaacaacacc tgaaaaagaa ctggtagcag cacaagtaaa acaaggcgcc ccagcgagcc 840

cacaaaaaac accaatcgaa caatgccgga agaagagaag cacaggacgt gaaagactaa 900

tgccgcaact cgaacacgaa gagaagccga actgcaacct accgacgaag tgcgatgaaa 960

tacgacagga agccagccgc cgcgcatagg caacatcttc gctagaatga cgacgagtat 1020

cgacataggt cgagagacga cgacgtagac accgaaccgt cgatacagag cgtt 1074

<210> 97
<211> 832
<212> DNA
<213> Homo sapien

<400> 97
aaaagggatg atgattcacc tatagggcga tggttctcta gatcatgctc gagcggcgcc 60

agtgtgatgg atgccgcccg ggcaggtacc tgggaggcgg aggttgcagt gagctgagat 120

cgtgacaccg cactccagcc tgggcgacag agcgagactc catctcaaaa aaaaaacaga 180

aacagaaaca gaaaaaaaaa aaaaaaggga ggaggcagag ccagacctca ttttacaaac 240

gcctgaagct gggggtaatc atggtcatag cgtgtccctg ggtggtgaaa ttggttattc 300

cggcgctcac aaatttccac cacaacaatt accggaggca agcgggagga gagtgaaaaa 360

cgaatgatag ggagacaaaa aaagaagagg aaaagaacaa gcaaggagga gaaagagaga 420

gaaaccaaac aaaagaagag acgagagaaa gcaaaagaag aggaaagcag agaaaggaaa 480

gaaagaaaaa aaagagatga ggaaagaaag caaagaaaga ggaaaacgac aaacaggaaa 540

cataggcaac agaaaagaga acaacgaagc aaaaccacac agaaagaaaa taagaagaca 600

aacaagggaa gagaacaagg caaagaagga acaagaacaa gagaataagg aaaaaaaaac 660

aaaagagcaa aggaaaataa gagagaaaac aacaacaaaa aaaaacaaaa gaaaaagaac 720

gcacaagaga aaaaaaacta gagaaaaaca aaagaagaca aaaacaaaac gcaacaagaa 780

ccaaagaaaa gagcacacga gaacaaagca cgaaaacacc aaaaggaacc aa 832

<210> 98
<211> 577
<212> DNA
<213> Homo sapien

<400> 98
gacagtaaag acgcaaggcc cggcgagatt gtcacaacat gcagatgaaa agaactcaga 60

gataaaaagc aattagcgac atcaaaagca cagacaaacc aagcacggaa aagcactgaa 120

gaagaccaag gctgaaataa gacagaacgt cagacacaaa agacagcgag agaagaacga 180

gggaaaggca gtactggaga gcaacaacaa cacagacaca ccaccagaca ccagcatgag 240

actcgaaaga agaaacgaga cggacacagg caagcgagca caaagccagg aaccaggaac 300

caacgacaga aacgagcggg gaaagaagaa gccggaagag ttcccaagcg aagagacagt 360

caacgggcga gtaagcgagc caagagaaac caggaagcaa atcggtcgaa gcaaacacac 420

aggggaccga gagacaaaga cgggaggcaa aggaaaaaag gaaagccaaa gaaggcagac 480

aggcaagaag agggaagata cagaaccaca tagggggccc aagaccacac aacaggcaca 540

aagcaagtac agaacgaaca gtaagagagc aacgaaa 577

<210> 99
<211> 1717
<212> DNA
<213> Homo sapien

<400> 99
cgtggtcgcg gccgaggtct ttttttttt ttttttctt tgttttttgg gggtgttgcc 60

ccaaattttt tggagaaata atagagacac atagggaaaa aaatttcaag aagggtctcc 120

cgcggagggg tttgggagaa tcactacctc ccaggtgtgt gtggggcgtg ggcgtgcacc 180

gtgagagggt gggccgagga ctgtgccgga agactcacga aaagctgtgg tgcgttatct 240

cggtggcgca atacgcgcgt ggtgtccgcg tggtggtgag agagtgtgtg tatatctcgc 300

ggcctcacca attctccacc accagcaatt aaccagaaca caatatccgg ctcaacaatt 360

ccacaccaac atacgaagca gaagaaaacc aatacaaaca tgacactgag cgagatagcg 420

agacaaagaa cggagccaca gaaatatgac caaaaaagag agaacaaaaa cccacagaga 480

cagacagaca aagaaagaac aaaaatgaac aagaaaacaa agaaaaaaaa aaacacaaga 540

cgagagagaa aaaaggagac gacgagaaaa acaagaaata aagagagaag cgagacaaac 600

agaacaaagg agcaacaaaa acaaaacgag caaaaaaaca acggaaccac tacaccacca 660

agaaaaccca agcaaagaaa acagaaacga gcgccattaa gcagacacac gaacagagag 720

agaaaaacaa aagacacaaa caaccaaaac acacatatag taggatagat agaagtgtaa 780

taataaaagg caggagtgaa gtataaacga tactccaagc cgaacaagaa gacatatcaa 840

aaagggagta gcataaatat cataaaccaa ataatacaga atgaatatag agtgcaagac 900

acaaacatct gatttagtac aacatagaaa aaaatatagg gatgaacaat aagagatcac 960

aaaaaaagag ggaagacaac tgctaaggaa aacagaccac aggatgagta aaaataaaaa 1020

gaggaacaca cgaaaagaga aaacagatga gagaaaaata acgagtagga aaacacaaga 1080

aaaaaaagga aaaacaacag caggagaaac ccaccggaca agagaaaaca aaaggcaata 1140

agcaaacaaa aaagaacaaa aacaaacaag caaaacgaac agccaaaaag aagagaaacc 1200

agaaggggga gtggggagag caaaacacaa ccaaaagaca aagaaggaga gaacaaaaat 1260

aaaccaagaa gaaaaaaaga gaaagaaaag agaaagacac acacacccac gggaaaacga 1320

gaagcacacg ccacacaaag aaaagaaagc aaacaaaaac aaaggatcaa taccaacaaa 1380

aaccagtgag cacgggagta gttagccaag ggaagaaatc aatcagaaca acacaggacg 1440

gcaaatacaa caaccccaaa ctctgtgaaa aaaacatgaa ggctagacac gaaatcaaca 1500

aacaaaaaca caaacagacg agatagaaag aaaaagagaa caataaatca gagacagaga 1560

aggaagaata agagcaatac gacaagagca cagcaagaac ataggtagga acagatgaag 1620

gaggcgaaga gaacaccaca gcagagcaga acacgacgcg gcaaaacgca aggaagagaa 1680

caacataaga aacggcaaac aaacaaacaa ggacaaa 1717

<210> 100
<211> 1423
<212> DNA
<213> Homo sapien

<400> 100
gtcgcggccg aggtacaagc ctttttttt ttttttttt ttttttttt ttttttcccc 60

ccgcgggtga ttttttttt ggtttttct ctatttgtga gtgtgcttgg gtgtgtggtg 120

caggtgcttt agtgtggcac aagccacttc tctctcgagg ggacccctcc caagaacccg 180

tggggtgtgg aattacggct ctgtggcacc cattcacgag gaaagcttgt gtgggttact 240

ctcctgtggt ctcacatatg ctgtgtgtct ctcctgtgtg tgtgtagaaa tgtgtgatat 300

atctcgcgct ctcacacata tctctccaca cacacacaca catatatgca ggagacacac 360

acagtgaaaa taagggagaa ccaaacataa aaacaaaaaa agaacggcca gcgaggagaa 420

cacacacaaa caaaaacata caaaaaacac aacacacaca agcaaaaaaa agaagaagaa 480

agaaaacaaa acaaaaaaaa aaacacatac ccacaaaaac caaacaacac aaacacaaaa 540

aaacacacaa aaaaaaaaaa aacaaaacaa accaacaaca aacaacaaaa aaaaaaaaac 600

aaaataacag acaaaaaatc aaaaataata aacacaacta taatatcact ataaacataa 660

aaaaataaaa caaacaaaca actaaaaaca aaactacaaa agacaccttа ccacacacat 720

ctccacacaa cacaacacaa cacaatacca taatatacga aaaacaataa ataaataaaa 780

aataatcaat aacacaccat cacatcaata acacaataaa caaaaatata ataatacata 840

tataaatcac acaaatactc actatatcta atactaaaca aaatacaaaa aaaaacaaaa 900

aaaacagaac acaaacatat caaacaaacg aaatcacaac acaaccaaat accacaccaa 960

actaaacaaa aacaaaaaaa accacacaac acaaaaaaga aaaaaataaa aacaaacaat 1020

caaacaaaac aacaacaaaa aaacaaacaa aaataacaaa aaaaatcata aaaaacaaaa 1080

aaaacacata cactattaat aataaaaaac aaacaaacaa caacaaacaa cacaacaata 1140

gacataccac aaataaaaac aaacaaacta ttataacaca gaacaacaac taacaaaaaa 1200

aaataatcaa aacataaaaa atataaaaaa aactaaaaat acacaaaaat ataaaatact 1260

aaacaaacaa aaaaaaatac aacaaaaaaa acaacaccaa aacataaaaa aaaacaaaac 1320

aaacacacaa aaaaaaaaaa aaaataaaaa aaaacataaa cacataaaaa aataaaaaaa 1380

aaaaaaaaaa aaataacacc aaaaaaataa acaacaaaac caa 1423

<210> 101
<211> 1627
<212> DNA
<213> Homo sapien

<400> 101
ttttttttt ttttttttt tttggttggt caatgaggat tattaattgg gggttaacat 60

aattgcaggg aaaaagggtc gggatagact ggggaagtgg ggagaagaaa cacccctctc 120

cccggggaac tcctgtgcag ctctccgagg ctctcccgtg gggtgtgggg tataaagtgt 180

ggggacacct tatgtgacac ttctcgtggt gagggcgccc accgtgtgtc ttctctctcc 240

gagcgggttg cgctctccct ctctcgagtg cgcgtggtga cacgcgtgtg cgcgagcgct 300

gtgtgtagcc tctctctgtg tgggtgagac acttctccca gctgtggctg ctgcgcgggc 360

gtctcgagtg actctcgagt gtgttagcgc tgtggctgtg gccgcggcgt tcacgtgtgt 420

gtgagtgtgt cgctctctgt gtgtgtgtgc gagagagcgt gtgtgtggtg tggcgtctct 480

ccacactctc ctcgcgtctc tgttgaagag agtgagcgag ttgtgtcgca cactctgtgt 540

cgctctctcc agcagagctg ccccagctgt gtcgtcacac agagtctctc cccgcgggtg 600

atacaagagt ctcccactgt ggagactcca cagacacaca cacactacag agtgtgtgtg 660

ggcccctctg tgtgtgtgcg gctgtgtgga gagactctct ctctcagaga gagagagaga 720

ggggggggagg acacacaaca gagagtgtga ccacagagtg tggggcgggt gtgcgccctc 780

tgagggggcc gagtaggcac ccctcatatg aggcagcggg gcgtagacac ccctgctggt 840

ggtgtctccc cccaaggtat cctccccgag aaaaaacaca ataatatgat gtaggatcat 900

cacacttata accttatatg cgggggggtc cccccacaaa cacagcggca gaaacagatg 960

tataaaatat aagactccga ggggcgccca taacaactcc gccgcggggg gtatatcacg 1020

aaagcaccac acaagcgttg atatgtgggc ctcaccacgg ggggttggag gcaccgccgg 1080

ggtgttgttc ccccccccac aagaagctgt tgtggcggtc atcctccccc aaagaaagaa 1140

aacatttccg ggcacaacga ggggggagaa ctcccccccа tgagaagggg ggggccgccg 1200

aggagatagc cgcagataaa taccaaactc tcaaatgaga atgaaaatta gtcaaccacc 1260

agaaatggcc acccacatgg tgtgtgtggg cctccctgtt tggggacccc attaggtagc 1320

gacaactcat cgtggtggtg tggtgcgcac cccccaatcg gtgtggccga caccaatctc 1380

cttcttctat tatatcttct tccccccaaa ctatatggaa aaacgcctgt ctgcagcggg 1440

gtcagaaaaa accatcattc atgtggggcg gccacaaata aaaaccatca ggtgtcgatc 1500

acccccttgg gtggttgtgg caagaaccaa ctttctgtcg gtcaaatccc cccgagctgc 1560

aaccacaaat tttcgcccca ccattacatc atcatcaaaa tcaaacagca atcaaacaag 1620

aaactcg 1627

<210> 102
<211> 936
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (401)..(401)
<223> a, c, g or t

<400> 102
gcgtgggtcg cgggcgaggt acaaggcaga ggccacacgg aggattcttg acactaagaa 60

gcacagggga caagcagaat cgttaatact gcgctgtgtg ctttggtgtg atcgtcagaa 120

gttgtgttgg acagacactg tgtccaggtc aggaaactga tgctataaca gacaagcatg 180

ttcacacgat acagcattta gaacacaggg tgagcatagt cggctcgtgt caaggaggag 240

taactcgcta tggaggagag tactaggcag aagaggagca gagaattgcc tccgagttag 300

cagtagtgta tagactgttg gtgtatgcta caccggtgca aggcagtgct acaacattgt 360

cctggagtta agagctgtac ctaaatagag cgcggttctg nttatctctt aagatagaca 420

catattctat agacaccatg atacctttga tctgcggcag tagacatagc gtcactagag 480

aagtaagaga ctttttgaag ctgcatgcac gagtatcggt caaagcgtat ggactgggtg 540

aggggagca tgagtaggag tttccctact tctagcgtgt gtcgaaagtc agaagaccga 600

cttgagagag tacatgggac agataaatat aacgtgtgtg gtcgtgtggt atattgcgtg 660

aacgggggg ttttcgccgt aaaccgaatt tgtctgctta gtttagcaga gggaagatcg 720

gaatttccac tgggcgtcag tgtccgatat tggtttaaag acttgaacat cgagaaccga 780

atttgtttct gtagggcgta aagcacaggg gctgggttcc agcagagtta gtaccagacc 840

atttcgtttt taccttgaca aaaagtcgaa ggagataaag ccttgggcgt tgaaggtcca 900

tgggtgccat tagcctggtt accctggtgg tgaaaa 936

<210> 103
<211> 502
<212> DNA
<213> Homo sapien

<400> 103
gcggccgccc gggcaggtct gtaatcccaa cactttggga ggctgaggca ggaggattct 60

ggggtcagga gttcgaggcc acactggcca acatggggaa agcccgtctc tactgaaaac 120

acaaaaatta gctgggcatg gtggtgggtg ccggtaatcc cagctactca ggaggctgag 180

gcaggagaat cgcttgatcc caggaggcag aagttgcagt gagctgagat cgcgccattg 240

tactccagcc tgggcgacag agcgcgactc catctcataa agaaaaatat tttaaaacca 300

tttctaaaac aaaaaaaaaa aaaaaaagaa aaaaaaggct tgggggtacc ccgtgtggcc 360

aaatagcgtg ttccctgtgt gttgacagtt gtgttttctc cgctccaaca aattctccca 420

ccaccaacaa tatacgacga caaaaagggg cgagcggagg agcgggcccc gaaccggcga 480

ccgggaaccc gcgcgcgaga ac 502

<210> 104
<211> 702
<212> DNA
<213> Homo sapien

<400> 104
gcgtggtcgc ggccgaggta ccagcccaga acccagtagc tcttctgggt ggctagaccc 60

agaagagcaa taacaatcac agcagttggc tctcgggaag cccatcccta ggggaagggg 120

gagaacacca cattaaggga tcaccctgtg gaacaagaga atctgaacag cagctcttga 180

gcttcagatc tttcctctaa cgtagtctac ccaagtgaga aggaaccaga aaaacaattc 240

tgataatgac aaaacaaagt tctataacac ccccaaaaga tcacgttagc tcaccagcaa 300

tggatccaaa ccaagaagaa atctctgaat tgccagaaaa agaattcaga aggccaatta 360

ttcagctact caaggagaca ccagataaag gtgtaaacca acttaaagga attaaaataa 420

taatacagga tatggatgaa aaagtctcca gagaaataga tatcataaat aaaaatcaat 480

cacaacttct ggaagtgaaa gacatactta gagaaataca aaatacactg gcaagtttca 540

acaatggact agaacaagta gaagaatgaa ctacagaact cgaagacaag gctctggaat 600

taacccaatc cagcaaaaaa aaaacaaaaa aaaaaaaggc tgggggaaac cggggccaag 660

gcggcccggg gggaatggtt ccggccacaa tccccaacgc aa 702

<210> 105
<211> 433
<212> DNA
<213> Homo sapien

<400> 105
aagatgatga atatataggc gaatgggcct ctaatgcatg ctcgagcggc ggcagtgtga 60

tggatggtcg cggcgaggta ctcgagcctg ggtgacagaa tgagactcta tctcaattaa 120

aaaaaaaaaa aaaagagtaa aacatctcca taccttaaaa aaaaattctg agcctctatt 180

tttagacttg tgatgattcg atacgaccaa tgtattttat cattgttatt ttaattatta 240

tttgctcttg ccaaagcacg ttctgtgatt tggtgcttct agtttgcttg ttttcatttt 300

aagaaccaga cacttctctc aaatcctttt tttaaagatg gaggtataga taagtgaatt 360

taaagaaaca ggtaaaaaat aataatttag tgttctggat tcttcttaac agaactttac 420

agactagcat ggc 433

<210> 106
<211> 2667
<212> DNA
<213> Homo sapien

<400> 106
ctgctgaagc tgctgcaggt gctgattgtc ttggaacacc acctgggtcg ggcccatgag 60

gaggcggaaa accagcccga cctgtcccgg gagtggcaga gagccctgaa cttccagcag 120

gccatcagcg ccctgcagta cgtgcagccc caccccctca cctcccaggg tcttctggtc 180

tctgcggtgg tgaggggtct gcagcccgcc tacggttacg gcatgcatcc ggcctgggtg 240

agcttggtca cgcattcctt gccctacttc ggaaagtccc tgggctggac ggtgacaccc 300

tttgttgtcc agatttgcaa aaacttggat gacttggtca agcagtatga aagcgaatct 360

gtgaagctct ctgtcagcac aacctccaag agggaaaaca tttctccaga ttatccactc 420

acccttctag aaggtctaac gaccattagt cattttgtc ttttggaaca agccaaccaa 480

aacaaaaaga ccatggctgc aggtgatcct gccaacttga ggaatgccag aaatgccatt 540

ttggaagagc tgcctcgaac tgttaacacc atggcccttc tctggaatgt tctcagaaag 600

gaggagactc aaaagagacc tgtcgatctc ctaggggcca cgaagggatc ctcttccgtt 660

tactttaaaa ccaccaaaac cataagacaa aaaattttag acttcttaaa ccccttgacg 720

gcccatcttg gggttcagtt gacagcggct gttgcggcag tgtggagcag aaagaaagcc 780

cagcgtcaca gtaagatgaa gattatccca acggcaagtg catcccagct aacccttgtc 840

gacttggtgt gtgcactcag caccctgcag actgacacgc tgctgcacct ggtgaaggag 900

gtggtgaaga ggccacccca agtcaaaggg ggtgatgaga aatcgccccc agtggacatt 960

cctgtgttgc agttttgcta tgcttttctc caaaggtaat acagtccccc gtcctccaaa 1020

aacttcctta aatacagatg ctattgcagt gagcatgcat aataaatatc tggttttatt 1080

tctagactac tagagagcca ttgttcagaa aatatcttaa agttgtcata atttcttccc 1140

aaggtattga tctatgcttt tccttctcca gagagttaac atcttaaaat ctgtgcagcg 1200

tactttgaac actttatatg agcgaagctt tatgtgaggc ctgttaaact ttaaaaggct 1260

tggatttgca ttaaattgat acagaaaaag aaaaaaacca catataggga agtgttaaag 1320

acctctttta gaaaaaggag aaatgggcca ggcgcggtgg ctcacacctg caatcccagc 1380

actttgggag gtggaggcag gcggatcact tgaggccagg agtttgagac cagcctagtc 1440

aacatggtga aaccctatct ctactaaaaa tacaaaaatt agtccggcat ggtggcgtgc 1500

gcctgtaatc ccagctactc gggaggctga ggcaggagaa ttgcatgaac ccaggaggcg 1560

gaggttgcag cgagccaaga cctcaccact gctctccagc ctgggtgaca gagcaaggct 1620

ctgtctcaaa aaaaaaaaaa aaaaaaaggc aaaatgattg tttctggtgg cgttctcagt 1680

gtgccttccc atgttttatg tggagaggta tctgctttga tttgctaagt taatatatgt 1740

attaggtgtc tcagctaaga gagcattaaa ggggattccg caggtttttc cccatggatg 1800

aaaaagaagc tttactggac ctcattcaga tcttacaatg gccgcatccc aggtgcagct 1860

ctgtgttgca gaaaatgagg gtgaggtggc tgggtgcggt ggctcagcgc ctgtaatccc 1920

agcactttgg gaggctagag tgggtgagtc acttgaggcc aggagtttga gaccagcctg 1980

gccaacgtgg taaaacccta tgtctactaa aactacaaaa aattagccag ggcttggtgg 2040

cagcatgcct gtagtcccag ctactcggga agctgaggca ggtgaatcac ttgaacccag 2100

gaggcagagg ttgcagtgag ctgagatcac accactgcac tccaggctgg gggacagaat 2160

gagactctat ctcaattaaa aaaaaaaaaa aaagagtaaa acatctccat accttaaaaa 2220

aaaattctga gcctctattt ttagacttgt gatgattcga tacgaccaat gtattttatc 2280

attgttattt taattattat ttgctcttgc caaagcacgt tctgtgattt ggtgcttcta 2340

gtttgcttgt tttcatttta agaaccagac acttctctca aatccttttt ttaaagatgg 2400

aggtatagat aagtgaattt aaagaaacag gtaaaaaata ataattagtg ttctggattc 2460

ttcttaacag aactttacag actagcatgg caaagcttct ctccgatctt agtgtggaca 2520

gtgctcgctg caaccatggg aataacctta ccaaatcact cttgaacatt catgataaac 2580

aacttcaaca tgacccagct cctgctcaca cttccataat gagctatcta aataagttag 2640

aaacaaatta cagttttaca cattcag 2667

<210> 107
<211> 718
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (611)..(611)
<223> a, c, g or t

<400> 107
agtgagggga ggtcagcgtg agggggcggg tggagaagaa gtgtccctac gaatgtcata 60

ggtctcagcc tcacccccac cacgggagac atagagctgc aggatcccag gggacggggt 120

ctcatccgtc ccaaccacaa gggcaatcaa agcccttctc cctgcgactc aataacaacc 180

gtcaaataaa aaatatcatc aatgacaatc aaaagaaaaa aagaaaaaaa aaaaaaaaaa 240

aagaatgagg gaaaaaaaac aaggaaaaaa aacaagaaga acacacggag gagagagaga 300

agagaaaaaa cggagacaaa gagacacaaa cgacacaaca gagacacgag agcacgaaac 360

accggacgca gcaacaaaga acacgcagaa acaagacaaa cgaacacaac agcgcgagca 420

caggaacaag aagaccagaa gagcaaggaa gacgagctag cggccaggca gacgaagaga 480

caggaggcca gagaagcaca caacacaggc gaaggagaag aagcaggacg gagaacgggg 540

aaaccgagga gagaaggaac gagagcagaa cagaaagaaa aaccaaagac agagacagca 600

gagccaaagc nagaagagga acgaagaaga gcgaacgacg acgaacacgc gcgcagaccg 660

caaggaagag aacggaacaa gagaagcagc agagaaacga gaaacagaag agagaagg 718

<210> 108
<211> 2112
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (2005)..(2005)
<223> a, c, g or t

<400> 108
atggaggtta agagtaatgg cagcgacagc aaggggaaca ggatccccag tcacttattg 60

ggctccagcc ctggggctga atgctactta attatcttca acctccaatc tgaggatgag 120

gccgagtatc actgtggaga gagccacatg tttgatggtg aggatggctc gagactgact 180

ctgactcatg gggcagctcc tgtgcgcagg ggagtctcag tctctgaggc ctcctatgag 240

ctgacacagc caccctcggt gtcagtgtcc ccaggacaaa cggccaggat cacctgctct 300

ggagatgcat tgccaaaaaa atatgcttat tggtaccagc agaagtcagg ccaggccct 360

gtgctggtca tctatgagga cagcaaacga ccctccggga tccctgagag attctctggc 420

tccagctcag ggacaatggc caccttgact atcagtgggg cccaggtgga ggatgaagct 480

gactactact gttactcaac agacagcagt gatgacatga accaggtaca ctgctctaag 540

ttctgcctta aggagagtgt tcctccacca ctgctgttca gggaagccca gaggccaggc 600

cacaaactag cgaacatggc caccctgacc atcagcaggg ctcagactga ggacgaggct 660

gactattact gtcacaggat aaagctggtg aaagagggcc tggatgaaag gacacacaaa 720

gcgtatcttt catctagtgg taaaggatgt gagttccata tggtgaagcc tgggtcaccc 780

cttggcccag acgtcctagg gtcctgggcc cagtctgtgc tgacgcagcc gccctcagtg 840

tctggggccc cagggcagag ggtcaccatc tcctgcactg ggagcagctc caacatcggg 900

gcaggttatg actatgtaca ctggtaccag cagcttccag gaacagcccc caaactcatg 960

atttatgagg tcgctaagcg accctcaggg gtttctgatc gcttctctgg ctccaagtct 1020

ggcaacacgg cctccctgac catctctggg ctccaggctg aggacgaggc tgattattac 1080

tgctgctcat atgcaggcag ctacacttgg gtgttcggcg gagggaccaa gctgaccgtc 1140

ctaggtcagc ccaaggctgc cccctcggtc actctgttcc cgccctcctc tgaggagctt 1200

caagccaaca aggccacact ggtgtgtctc ataagtgact tctacccggg agccgtgaca 1260

gtggcctgga aggcagatag cagccccgtc aaggcgggag tggagaccac cacaccctcc 1320

aaacaaagca acaacaagta cgcggccagc agctacctga gcctgacgcc tgagcagtgg 1380

aagtcccaca aaagctacag ctgccaggtc acgcatgaag ggagcaccgt ggagaagaca 1440

gtggccccta cagaatgttc ataggttctc atccctcacc ccccaccacg ggagactaga 1500

gctgcaggat cccaggggag gggtctctcc tcccacccca aggcatcaag cccttctccc 1560

tgcactcaat aacaaccctc aataaaatat tctcattgtc aatcaaaaaa aaaaaaaaaa 1620

aaaaaaaaaa aaaaaagaat gagggaaaaa aaacaaggaa aaaaaacaag aagaacacac 1680

ggaggagaga gagaagagaa aaaacggaga caaagagaca caaacgacac aacagagaca 1740

cgagagcacg aaacaccgga cgcagcaaca aagaacacgc agaaacaaga caaacgaaca 1800

caacagcgcg agcacaggaa caagaagacc agaagagcaa ggaagacgag ctagcggcca 1860

ggcagacgaa gagacaggag gccagagaag cacacaacac aggcgaagga gaagaagcag 1920

gacggagaac ggggaaaccg aggagagaag gaacgagagc agaacagaaa gaaaaaccaa 1980

agacagagac agcagagcca aagcnagaag aggaacgaag aagagcgaac gacgacgaac 2040

acgcgcgcag accgcaagga agagaacgga acaagagaag cagcagagaa acgagaaaca 2100

gaagagagaa gg 2112

<210> 109

<211> 2168
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1144)..(1144)
<223> a, c, g or t

<400> 109
agccccccgg ccgcgggtaa tgacactata ggcgacttgg gctctctaga tacatactcg 60

agctggcgcc gaggtataat aggatatgcc tgaggtacga gaagcacagt ccctaaattt 120

ctagctagct ataggaacca gataagaaat gaagaaaaaa gaaggcatat caatgataca 180

agatacacgt atctaagtga ggagtgagga ctaccctaca tactatacta agacacagtg 240

cggtacaaga agcatgatac gatgactgtg cgtgtcacat atactaatgt actaagtgag 300

gtacggcgac cgataccaaa atatgcccca atgtgcctgg tgctccacag catcttacca 360

tatcccatgc atgcaaaatg catggtaagc acatggtgtc caaatgtgtc agcctactat 420

actaaaacaa ccacatgcag cacccataac agatgcaaca tgcaaagcac caaacaggga 480

cacacagcac aactcgctat cttaacgata gaacagatcc aatccccaga ctataacatg 540

ttattaaccc atggcctact acaggccgct caatggaacc tgggtttatc cttaaagcaa 600

caacgttatg cccaactcgc ctcaaggaca cgccacgcca atggcatccc ggcaaccgga 660

gcacgctcga gcaataatca cgaacatcgt cctgaacggc gggcactgcg ttaagtgctc 720

cgcgcacctc aagactgggc aatactctac ccccttaaa ctacatgaca acatggcccc 780

cttacctgct ggctaccctt accaccttaa caaggaccaa cagggccaga ccacacttta 840

tatggttacc gccccgggcc ttaacagaag tcccaaataa ctcctccccc gttacaagct 900

gcgccacagc gcaaccatat ataaaaacac caaccatggt gtgcacactc atacagcgcc 960

acaatggcac acacatggca attacacgag acggcatccg tagctatatg ctaacacact 1020

gccacagagg tttacatcgt gggtggcatt tatgcgcacc acacggccta gcgctaaagg 1080

ccatgtatat ttacgtaaaa cccgcaacaa gctgccccca ccagcgtcac cagcgccacc 1140

acanccatca atcggcaagc tttcacaaac tatcgccaat ttaacacatg ggtggccaac 1200

accaccgcgg ctggacaccc cccaggttgc cgtgcacata tcactgccac taactaccga 1260

ttatccccca cgctccagtc ctagttctct atgcacttat agccgaacgt cgagttccgt 1320

atataagcga gaacgaaggt ggttactaca accagccgct ggggtgcgca ctactcctat 1380

gacctcttag gtcagaccgt acagtgcgtg cagcagccca ccggcgtggg cggcctacca 1440

tattggcagc gcacgtctta tgaacaggca caccaatact ggtagtgtca aaactacgac 1500

atcggacaca cgcaacgtgt caacacgtta acaaacatcc cccgactcct gtaatactgc 1560

gacacaacta gtcgagtgtg gtaattctcc ccaccctcta tacatatcag aacccaccag 1620

cagcgtggca cccaccatgt tatctccgat gtgtatcggt tcagcacact gcggtggaaa 1680

ccccttgtgc tgcgaccttc cctacttaca ggcgtcgtgt caactttccc cgggcggtat 1740

ggcacattag tgccgcgcct tagactaaca cttccacttt gtcgcgctgc catcgacacc 1800

tttgctgctt tatttttcgc cctctttgac ttctgggtca gtaaaatatt gcccatccga 1860

ttctaggtgc gtgatgcata ccatagcgat tagtataaat atcccattac ggatcaaagg 1920

cgttgacatt accccgtatt gtgtgtgcta tgaccgtccc ataacgaggc gggtgctacg 1980

tatcgggtac tcagctcttc ataacgcccc taataaatac tatatatcac ggggctccat 2040

acagggtatt actacacgag tggtgtacta atacagcgcg ctctcgtgtt gatctgctcc 2100

tagtaatacg gtgaaaatac cactaaacta accggccttc ggcggataga cacatgtcgt 2160

gataagcg 2168

<210> 110
<211> 959
<212> DNA
<213> Homo sapien

<400> 110
ggggccgggg cgggcggtat acttcaagat atacaaaaga ttatcccagg gcatactggc 60

tgagatagcc tgcgcgaatc actagtccac tacgggtgtg gcgagattgg agacgaagtc 120

tgctcggctt agagtccacg tggagtgtct tggcaggggt ttgcacacgg acggaccgcg 180

agaaaaggta aacccttcgg agcctccaca ctccgggggt gataagccgt ggatcctctc 240

ggtgatgaca ctaagcttat actcaccgac atacaacaat atgcacgcaa ccaagaccta 300

attgctgtat gtgagggttc cttacagcga aggggcaaga gcgtttgtgg cgtacagtcg 360

agtaggtcga ataagcttgt tttccgttgt tgtttgaaga atatgttaat accgcttcaa 420

cagattttcc cctaagccaa cgaaaagcct attaccgcgg gaaatggcca aactctagga 480

gggaccgcgt gaggtccctt tacccgcctt gggcattccc cagatgggtt atgggttgag 540

ccaccgcgcc gttgtgggcc cccaccgggt cgtccgcccg ttatatccac gcgtaaccag 600

agggcttaat ttaccgggga aaccctcccc acgataacgt ccgtttaact tgggggggcg 660

cgcttaccta tggattagcg gtcgaaggtg acaataagga gaaccaatac cggttcgaag 720

aaaaacgcgc gcatttaggt tgccgttgat atgaagagac ctctcatacc agagcgcgag 780

actccccaat atcaaacgag ccacgttggg ttgtatcacc cgaccaatcc gatatatgac 840

ttatgacaag cagacaatta taaggttaag atatattcgg cacgcagggt tcacatacca 900

aacccaaaca gactatattc gcacacaaga ggaggggccg cattccccca tgtgatatg 959

<210> 111
<211> 815
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (206)..(305)
<223> a, c, g or t

<400> 111
ggaatgatac actcactata ggaccattgg ttactctaga tgcatgctcg agcggcggta 60

tgtgatggat agcgtggtcg cggccgaggt acctaaacag gccaaatgtt gcctttgggg 120

ttcctgtttc aacagcatgg tgtgaagcgc tgcatcaacc ttctctgcct catcctgcaa 180

ggtggcaaat tcctcaagaa tatgannnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300

nnnnnagcgc ggggagctaa gggagtaaga ctaagtggaa agagaagtaa gaagtagaac 360

atgacgatgg agaggataac taaaagaaga gaaagagcat gaagtagaca agaattgaat 420

aaagatgagg catagaaaac gaaagagcac gagaagaaaa aagaggagaa gaatagaaga 480

gaggcatgtt acagagaata gagatcaaga gagatcaaaa gacaggccac aaagacaaga 540

cggaggagga gaacgaaaaa gaagtcagaa gaaaacaaaa aacgagagaa taacagaaat 600

caacacagca acaagagagc agacaaggca agagcaaaag aaacacaagc aacagagaga 660

agccaaacga aaaaaaagaa aagggagaca gcagacgaaa gagaccaagc gacaccgaca 720

gatggaacgc aaaagagaac agcacagaga ggaaggaagg aaagaaatcg aaaccagggc 780

gaagcgggcg agaaacaaga aaagagaggc caggg 815

<210> 112
<211> 736
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (439)..(439)
<223> a, c, g or t

<400> 112
gccgaggtag gagaatcgct tgaacacggg aggcggaggt tgcagtgagc cgagatcgcg 60

ccactgcact ccagcctgga agacagagtg agactccttc ccaaaaaaaa aaaaaccttt 120

aaaattggat ttggaagttg gattattctt ctcataattc ttctaattct ctccttttag 180

agatgtgatc cagctccatt taagacgact tggcagattg ccagaacctt attgcccttа 240

ttaaattcca ttaaatttaa ttctcagatt tatttggaga aggaaggtaa gattttctta 300

ttagaacccg cacacttgga acctgggtta agcgcttggg cggtaactca tgggctcata 360

ggctggttcc cgtgggtggt gaacattggc ttattccggc ttccacaatt ctcccactac 420

aacattccgg gaagcaacnt cactggaaga tgaataatgg cagatgtgtg aattggagca 480

acactctact tcattggact cagtggactc ctagatgcgc aaaacatcac aagaaggatg 540

ggggccagag atctacagat ggtcatcata caacgagaag cattacaagt gagaactatc 600

cacgaacgaa caaagagctg aaatgagata ctgaaggtca tatatgcacc ggataacgga 660

cagtagacaa tagactccct ttggagagat ctggaccaga gatggatatc aatgatatgg 720

caatatgctg gatcca 736

<210> 113
<211> 588
<212> DNA
<213> Homo sapien

<400> 113
tctcgactgc gcctatgtga tggatgccgc ccgggcaggt ccccccтttt ttttttttt 60

ttttttttt ttttaagag gggtttaaaa aatttttctt tttggaaaat tttctggaaa 120

gtatttaaaa ccccctttgg ggaaggaaaa aaaccaaccc aatgtgaaat tttaggaaaa 180

aaaagtgcga aaagcagcgt gcgaaaactc cgtgcgccct ttccacccca gggggcccac 240

gcccggaaat taacgcgtgg gggataacca gggccccata aggcgtgtgt tcccgcggtg 300

tgtgacaagt gtggatatct ccgcgcccac caattctccc caacaacaca ttcccgaaac 360

aaaacgggaa gagaggaaaa aaaaaaaaca aaaaaaaaaa aacagagtac aaatataaca 420

acgcaaacgc atactcgggg cccaagcgga ggtgaaggtc agaagaataa aaagagagaa 480

gcgagcgagc agcggtcgag cgagagaaaa gcagacacaa acaacagcca accaaggaag 540

ggagcagaag aaaacgaaag aggagaaaca aaggcaaaga aagacaaa 588

<210> 114
<211> 1098
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (327)..(327)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (342)..(342)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (471)..(471)
<223> a, c, g or t

<400> 114
ccggcccggg ccggtggcgc ttcgggagcc gcgggttatg tttgcagaca tggacaagtc 60

caatgaacca caccactatc acaaaccagg tcccgaagag atgggcacga gaaatgggta 120

ctcaagaaga ccccgattca tcgaatgagt actgacattc ttcgtctaca cgggcggttg 180

cgaccaccaa aggttccctg gaggagaaca tggcgcattc tgcctcgtca cggcacatgc 240

gaaggatggg ccttagcagg caagatgcac ggccgagagc gaagtgccga gaggccagtg 300

ggatgctcct gttgccggag tgcctantca atggcccgtc gntcaaagcg ccaatgggac 360

gggttaacgc cgcagagtat cccagatggg ctggtcatgc ccgcatatgc ctcgactcga 420

ctttggccgt acagttgctc cacgcccgaa gcttggctgt aagtcagtgg ntcatagcat 480

gtttccctgt gtgaaagttg ttagtccgct tcacaagttc catcacaaca taccggagca 540

tggcgcatct tgagtaacgg cctctgtgat gaggcttagc atcagctatt tgcgtgctga 600

ggaaaacata tctggactgc tggtgcatac cggcactatc gaaagactga cactgaaaag 660

caacagactg acatggccac aacactaccg gaacagccct agcgcatgcg ataaaggtat 720

catagggtat cgtcaacagc atgcatgcgt gaaccatgga tataccatat aactggaata 780

atggtgaaac acaatacaat ggggaattta actccagatc acgacactaa cctgggacac 840

cgaaggaata cggagatttt aactaccaat cacatggttg aacccataga aaaggcaaca 900

tgaagcaagc aagactggcc ataccaacca acacaggaaa cagggcgccc atggcggaac 960

aacaaagggg ccacaaccac agcacacacg acaaccaggc gcaccaccac gggccggtca 1020

taaaccacgg acccatacag caggccaccc cgcgtgaatc aacatggcaa tcaagggaca 1080

caacagacac acacaacc 1098

<210> 115
<211> 816
<212> DNA
<213> Homo sapien

<400> 115
gagaactagt ctcgagtttt tatttattgt tttttagggt gtttctcttt ttttggggaa 60

ccgcttcttg ctgtgtccgc ccaggcttga actgcatgtg ttgcgatctt gggcttcgtt 120

gcatcgttgt tgctctttct gggtttcagc ggggtgtcta gtggtccttc tacccctcct 180

tgtaaatgag ttagtgtttc cgtggttgtt attgtccccc cagcgcccgt gggtctattt 240

tttatcattc ttgtgttttc acgattaaca aaacagtgtt tttcccccct ctgttgggtc 300

ctggtctgtt ttccggaagc tccgtgcacg tctgtattac agcctcgcag agtctccaaa 360

cccactctcc aagtgcggca gcgtgaatta taggcgaggc tatgtgtagc acgcctacca 420

cggagccctg cacacagatg gtggttatct acccctcgtg tgcacaccat gtttttgtgg 480

cgcctcgctg agcttattgt ggttaacaga aggtgctctt ggtcgcaatt agtgtacaac 540

gcttggagct ctaacctttt ttgtgtggta acacccgtgg tattttgcat gtgaagagaa 600

cgggtccatt ataaaggcga gagaaaagta agacctgttt gtcactattt ctgtttccat 660

gtgtaaccgt tgtttttttc cccccaaaat taaccgactt tttttacttt tgcaaaaaaa 720

aaaaaaaaag gtcttggggt aaccacaggg ccaaacgggg tccccgggga aaatttttt 780

accgggacac aattccccaa tacttagaaa aaaaac 816

<210> 116
<211> 33
<212> PRT
<213> Homo sapien

<400> 116

Met Leu Val Ala Asp Phe Phe Phe Thr Gln Asn Lys Val Gly Arg Cys
1 5 10 15

Thr Cys His Val Glu Tyr Leu Lys Lys Thr Lys Cys Leu Phe Lys Arg
20 25 30

Glu

<210> 117
<211> 18
<212> PRT
<213> Homo sapien

<400> 117

Met Ile Leu Asp Ile Cys Leu Tyr Ala Ile Met Ala Tyr Val Met Ile
1 5 10 15

Met Asn

<210> 118
<211> 52
<212> PRT
<213> Homo sapien

<400> 118

Met Thr His Val Cys Ala Thr Ala Leu Gln Pro Gly Arg Gln Ser Glu
1 5 10 15

Thr Pro Ser Gln Lys Thr Lys Thr Lys Gln Asn Glu Thr Ile Asn Lys
20 25 30

Val Thr Asp Asn Leu Gln Asn Gly Arg Lys Tyr Leu Pro Thr Met His
35 40 45

Pro Thr Lys Ile
50

<210> 119
<211> 192
<212> PRT
<213> Homo sapien

<400> 119

Lys Ala Asn Asn Ala Gln Ser Asn Arg Gln Pro Thr Glu Trp Ala Lys
1 5 10 15

Ile Phe Ala Asn Tyr Ala Ser Asn Lys Asp Leu Ile Ser Arg Ile Tyr
20 25 30

Lys Lys Leu Gln Lys Ile Tyr Lys Arg Lys Thr Ser Asn Pro Leu Lys
35 40 45

Arg Lys Trp Ala Lys Asn Met Asn His Ile Ser Lys Glu Asp Ile Tyr
50 55 60

Ala Phe Lys Lys His Ile Lys Asn His Ser Ser Ser Leu Ile Thr Thr
65 70 75 80

Glu Val His Tyr His Leu Thr Pro Val Arg Met Ala Val Thr Arg Lys
85 90 95

Ser Ile Asn Asn Arg Cys Trp Gln Gly Cys Gly Glu Asn Gly Thr Ile
100 105 110

His Cys Trp Trp Glu Cys Lys Leu Val Ala Pro Leu Trp Lys Ala Gly
115 120 125

Trp Ala Phe Leu Lys Glu Leu Arg Ile Thr Ile Gln Leu Ser Asn Pro
130 135 140

Ile Ile Pro Lys Gly Met His Ile Pro Arg Lys Tyr Lys Ser Leu Tyr
145 150 155 160

His Lys Gly Thr Cys Thr Cys Met Ser Ile Ala Ala Leu Phe Thr Ile
165 170 175

Ala Lys Ile Arg Asn Gln Pro Lys Cys Ala Leu Ile Ile Gly Trp Leu
180 185 190

<210> 120
<211> 99
<212> PRT
<213> Homo sapien

<400> 120

Met Ser His Ile Cys Ile Tyr Thr Lys Lys Leu Gly Arg Arg Thr Tyr
1 5 10 15

Tyr Ser Pro Pro Thr Ser Gly Val Arg Gln Arg Gly Glu Arg Glu Gly
20 25 30

Thr Pro His Gln Arg Val Pro Thr Pro Gly Glu Asp Thr Glu Arg Ile
35 40 45

Pro Thr Pro Glu Asp Arg Gln Pro Arg Arg His Ile Tyr Val Gly His
50 55 60

Asn Lys Asp Thr Gln Glu Asn Ala His His Ser Ser Asn Tyr Ala Arg
65 70 75 80

Arg Arg Arg Arg Lys Lys Glu Pro Ser Gly Arg Thr Gly Glu Thr Asn
85 90 95

Leu Arg His

<210> 121
<211> 21
<212> PRT
<213> Homo sapien

<400> 121

Met Gly Gln Asn Trp Met Asp Leu Leu Lys Gly Asn Ile Glu Gln Asp
1 5 10 15

Asp Glu Leu Ser Lys
20

<210> 122
<211> 79
<212> PRT
<213> Homo sapien

<400> 122

Met Phe Leu Val Ser Ser Phe Asp Ile Val Leu Phe Ser Cys Leu Phe
1 5 10 15

Leu Arg Pro Leu Val Leu Cys Cys Pro Phe Ser Pro Ser Ser Tyr Val
20 25 30

Gly Leu Cys Gly Val Tyr Phe Pro Val Leu Phe Leu Thr Ile Arg Phe
35 40 45

Val Phe Phe Phe Phe Phe Val Ser Pro Phe Ser Cys Phe Leu Phe Leu
50 55 60

Arg Leu Cys Ser Ala Val Val Pro Leu Val Gly Ile Val Cys Leu
65 70 75

<210> 123
<211> 27
<212> PRT
<213> Homo sapien

<400> 123

Met Val Phe Lys Pro Val His Asn Thr Val Leu Gln Phe Ser Glu Leu
1 5 10 15

Pro Pro Thr Gly Ile Ile Ile Pro Gln Tyr Pro

20 25

<210> 124
<211> 54
<212> PRT
<213> Homo sapien

<400> 124

Met Phe Arg Pro Gly Phe Gly Tyr Tyr Ile Asn Pro Pro Gly Pro Pro
1 5 10 15

Pro Asn Pro Ala Ser Val Asn Arg Ala Asn Thr Leu Glu Asp Arg Asp
20 25 30

Lys Asn Phe Glu His Leu Phe Gly Gln Leu Leu Lys Glu Phe Leu Phe
35 40 45

Pro His Thr Ser Pro Gln
50

<210> 125
<211> 91
<212> PRT
<213> Homo sapien

<400> 125

Met Cys Phe Ser Val Thr Phe Ser Ser Ser Val Gly Leu Ser Phe Cys
1 5 10 15

Val Ile Ser Ser Phe Leu Leu Ser Cys Cys Ser Leu Ser Ser Trp Leu
20 25 30

Leu Ser Val Phe Ser Thr Arg Cys Cys Leu Glu Ser Val Gly Ser Gly
35 40 45

Leu Leu Leu Ala Phe Trp Thr Gly Pro Asp Thr Gln Leu His Pro Gly
50 55 60

Thr Ser Leu Trp Pro Arg Thr Thr Pro Arg Leu Leu Gln Glu Ala Leu
65 70 75 80

Pro Asn Leu Gln Val Asn Arg Phe Arg Asn Ser
85 90

<210> 126
<211> 53

<212> PRT
<213> Homo sapien

<400> 126

Met Leu Phe Lys Pro Leu Gly Lys Cys Ile Ser His Leu Thr Leu His
1 5 10 15

Glu Leu Leu Gln Gly Leu Gln Gly Leu Thr Leu Leu Pro Pro Gly Ser
20 25 30

Ser Glu Arg Pro Val Thr Val Val Leu Gln Asn Gln Val Thr Cys Leu
35 40 45

Gly Gly Phe Phe Pro
50

<210> 127
<211> 37
<212> PRT
<213> Homo sapien

<400> 127

Met Leu Leu Glu Arg Arg Ser Val Met Asp Trp Ser Arg Pro Arg Tyr
1 5 10 15

Phe Leu Tyr Pro Asp Ile Asn Leu Met Cys Cys Asn Leu Phe Asp Met
20 25 30

Ile Ser Tyr Lys Ile
35

<210> 128
<211> 50
<212> PRT
<213> Homo sapien

<400> 128

Met Tyr His Arg Glu Ile Val Pro Val Tyr Glu Val Leu Ser Val Ile
1 5 10 15

Thr Gly Leu Gln Ile Gln Val Phe Ser Gly Lys Glu Ala Asp Ser Val
20 25 30

Ile Lys Arg Ser Ile Gly Trp Gly Pro Phe Phe Lys Pro Arg Cys Tyr
35 40 45

Asn Pro
50

<210> 129
<211> 26
<212> PRT
<213> Homo sapien

<400> 129

Met Ala Arg Pro Gly Cys Arg Ile Pro Ile Gly Tyr Leu Pro Cys Ile
1 5 10 15

Ala Val Leu Phe Tyr Gly Phe Leu Val Leu
20 25

<210> 130
<211> 68
<212> PRT
<213> Homo sapien

<400> 130

Met Thr Ser Gln Gly Leu Ser Leu Leu Ser Gln Ser Gly Phe Phe Leu
1 5 10 15

Leu Phe Leu Ile Glu Ile Ser Leu Ala Leu Leu Pro Lys Leu Ser Arg
20 25 30

Thr Pro Gly Pro Gln Ala Ile Pro Arg Cys Pro Arg Ala Leu Pro Pro
35 40 45

Gln Ser Cys Trp Gly Leu Met Gly Val Ser His His Ser Gln Pro Gly
50 55 60

Lys Ser Val Ser
65

<210> 131
<211> 86
<212> PRT
<213> Homo sapien

<400> 131

Met Arg Met Trp Tyr Ser Arg Gly Thr Tyr Ser His His Ile Thr His
1 5 10 15

Leu Val Ala His Thr Pro Gln Glu Ala Ser Ala Phe Gly Arg Gly Gly
20 25 30

Ser Leu Ile Phe Tyr Lys Pro Val Gly Asp Ile Ser Arg Cys Gly Ala
35 40 45

His Ile Ser Ala Val Cys Ser Ala Val Val Cys Glu Asn Val Trp Tyr
50 55 60

Ile Ser Arg Leu Ser Pro Asn Ser Pro Pro His Lys Ile Arg Arg Thr
65 70 75 80

Thr Lys Lys Gly Gly Gly
85

<210> 132
<211> 111
<212> PRT
<213> Homo sapien

<400> 132

Met Ile Ser Gly Arg Glu Asn Val Lys Lys Asn Ile Asn Glu Ala Arg
1 5 10 15

Gly Gly Arg Arg Ile Lys Leu Arg Gly Gly Ser Thr Ile Glu Ala Pro
20 25 30

Lys Met Tyr Pro Ala Gly Val Val Ala Ala Pro Leu Phe Val Val Val
35 40 45

Ile Ser Pro Gly Leu Pro Thr His Ile Ser Pro Pro His Asn Gln Leu
50 55 60

Asp Arg Thr Gln Thr Thr Gln Asn Thr Thr Lys Gln Thr Thr Ser Lys
65 70 75 80

Lys Asp Glu Pro Asn Gln Arg His Arg Asn Thr Thr Asn His Lys Thr
85 90 95

Thr His Gln Gln Asn His Thr Thr Pro His Pro Tyr Arg Asn Lys
100 105 110

<210> 133
<211> 36
<212> PRT
<213> Homo sapien

<400> 133

Met Thr Phe Gln Gln Cys Ala His Thr Leu Ala Glu Ser Ile Trp Ile
1 5 10 15

Phe Ser Asp Val Gln Gly Phe Ala Thr Pro His Leu Phe Leu Arg Ser
20 25 30

Tyr Leu Ala Met
35

<210> 134
<211> 35
<212> PRT
<213> Homo sapien

<400> 134

Met Leu His Val Asn Arg Val Leu Cys Leu Val Ala Ser Pro Gly His
1 5 10 15

Glu Arg Gln Ser Glu Thr Leu Ser Gln Lys Gln Lys Lys Lys Phe Leu
20 25 30

Leu Leu Pro
35

<210> 135
<211> 94
<212> PRT
<213> Homo sapien

<400> 135

His Pro His Thr Arg Leu Asp Val Cys Val Cys Leu Cys Val Cys Met
1 5 10 15

Cys Val Cys Met Cys Val Glu Thr Gly Phe Arg His Val Ala Arg Val
20 25 30

Cys Val Cys Val Cys Val Cys Val Cys Val Cys Val Cys Arg Asp Trp
35 40 45

Val Ser Pro Cys Ala Gln Val Cys Ala Cys Val Cys Val Cys Val Cys
50 55 60

Val Gly Thr Gly Phe His His Val Ala Gln Val Cys Val Cys Val Cys
65 70 75 80

Arg Asp Trp Val Ser Pro Cys Cys Pro Gly Val Cys Val Cys
85 90

<210> 136
<211> 66
<212> PRT
<213> Homo sapien

<400> 136

Met Leu Val Gly Trp Phe Phe Val Phe Val Leu Val Cys Gly Glu Thr
1 5 10 15

Gly Phe Cys Cys Phe Pro Gly Tyr Ser Lys Val Leu Gly Ser Ala Cys
20 25 30

Ile Ser Leu Pro Gly Ser Trp Asp Tyr Arg Arg Glu Pro Leu Cys Pro
35 40 45

Ala Leu Arg Asn Asn Phe Leu His Leu His Ser Ser Asp Ser Trp Phe
50 55 60

Val Pro
65

<210> 137
<211> 137
<212> PRT
<213> Homo sapien

<400> 137

Met Asp Val Ala Asp Glu Val Ile Leu Val Ile Glu Leu Gln Lys Leu
1 5 10 15

Leu Val Asp Phe Phe Phe Phe Phe Phe Phe Phe Trp Lys Arg Phe Leu
20 25 30

Pro Leu Ser Pro Gly Trp Leu Arg Gly Cys Leu Gly Leu Asp Pro Arg
35 40 45

Pro Pro Gly Ala Val Ile Ser Leu Pro His Phe Pro Leu Leu Gly Leu
50 55 60

Arg Ala Cys Thr Thr Thr Pro Ser Tyr Phe Trp Tyr Phe Ile Ala Glu
65 70 75 80

Thr Gly Phe Pro Ser Val Gly Arg Ala Trp Phe Ser Asn Phe Pro Thr

85 90 95

Leu Lys Leu Thr Ser Ala Leu Leu Gly Pro Ser Gln Ser Cys Val Gly
100 105 110

Leu Pro Gly Val Glu Pro Arg Pro Trp Pro Pro Ile Phe Pro Leu Ser
115 120 125

Ile Asn Ser Asn Ser Trp Pro Ser Leu
130 135

<210> 138
<211> 61
<212> PRT
<213> Homo sapien

<400> 138

Met Asp His Glu Leu Pro Pro Asp Phe Ile Val Gly Gly Leu Pro Leu
1 5 10 15

Lys Lys Leu Gln Pro Thr Gln Pro Phe Tyr Lys Thr Cys Leu Val Leu
20 25 30

Pro Leu Arg Ser Phe Pro Ser Asn Leu Cys Phe Ser Pro Cys Ser Pro
35 40 45

Pro Tyr Glu Phe Ser Asn Phe Ser Ser Ser Ser Pro Val
50 55 60

<210> 139
<211> 41
<212> PRT
<213> Homo sapien

<400> 139

Met Pro Pro Gly Ile Phe Ser Pro Ser Phe Pro Phe Phe Ser Leu Ser
1 5 10 15

His Ser Glu Ala Val Gly Ser Phe Asp Glu His Ile Pro Ser Thr Gly
20 25 30

Gln Glu Ser Cys Cys Leu Ser Ile Trp
35 40

<210> 140
<211> 39

<212> PRT
<213> Homo sapien

<400> 140

Met Leu His Thr Ala Gly Cys Arg Asn Ala Ser Arg Gly Gly Ala Asp
1 5 10 15

Thr Phe Arg Val Asp Arg Glu Arg Gly Leu Pro His Thr Asp Ser Gly
20 25 30

Lys Ser Gln Gln Ser His Met
35

<210> 141
<211> 51
<212> PRT
<213> Homo sapien

<400> 141

Met Leu Pro Cys Arg Lys Ile Pro Ile Thr His His Val Ser Gln Cys
1 5 10 15

Cys Val Trp Arg Pro Gly Phe Val Pro Leu Pro Arg Ile Ala Val Ala
20 25 30

Asp Ile His Arg Asp Pro His Met Asp Val Cys Met Lys Ile Pro Leu
35 40 45

His Arg His
50

<210> 142
<211> 40
<212> PRT
<213> Homo sapien

<400> 142

Met Leu Ala Asp Leu Ala Leu Ser Ser Ala Thr Ser Ser Thr Pro Val
1 5 10 15

Ser Glu Ala Arg Asn Leu His Cys Ser Ser Glu Leu Pro Gln Asn Asp
20 25 30

Val Leu Leu Ser Lys Glu Asn Ser
35 40

<210> 143
<211> 192
<212> PRT
<213> Homo sapien

<400> 143

Pro Gln Lys Arg Lys Arg Gly Ala Glu Val Leu Thr Ala Gln Phe Val
1 5 10 15

Gln Lys Thr Lys Leu Asp Arg Lys Asn Gln Glu Ala Pro Ile Ser Lys
20 25 30

Asp Val Pro Val Pro Thr Asn Ala Lys Arg Ala Arg Lys Gln Glu Lys
35 40 45

Ser Pro Val Lys Thr Val Pro Arg Ala Lys Pro Pro Val Lys Lys Ser
50 55 60

Pro Gln Lys Gln Arg Val Asn Ile Val Lys Gly Asn Glu Asn Pro Arg
65 70 75 80

Asn Arg Lys Gln Leu Gln Pro Val Lys Gly Glu Leu Ala Ser Lys Leu
85 90 95

Gln Ser Glu Ile Ser Arg Gly Cys Gln Glu Asp Gly Ile Ser Ile Asn
100 105 110

Ser Val Gln Pro Glu Asn Thr Thr Ala Ala His Asn Asp Leu Pro Glu
115 120 125

Asn Ser Ile Val Asn Tyr Asp Ser Gln Ala Leu Asn Met Leu Ala Asp
130 135 140

Leu Ala Leu Ser Ser Ala Thr Ser Ser Thr Pro Val Ser Glu Ala Arg
145 150 155 160

Asn Leu His Cys Ser Ser Glu Leu Pro Gln Asn Asp Val Leu Leu Ser
165 170 175

Lys Glu Asn Ser Leu Arg Gly Thr Ser Asp His Glu Tyr His Arg Gly
180 185 190

<210> 144
<211> 24
<212> PRT
<213> Homo sapien

<400> 144

Met Leu Pro Leu Gly Phe Leu Phe Gln Gln His Gly Val Lys Arg Arg
1 5 10 15

Ile Asn Leu Leu Cys Leu Leu Lys
20

<210> 145
<211> 733
<212> PRT
<213> Homo sapien

<400> 145

Met Val Met Lys Ala Ser Val Asp Asp Asp Asp Ser Gly Trp Glu Leu
1 5 10 15

Ser Met Pro Glu Lys Met Glu Lys Ser Asn Thr Asn Trp Val Asp Ile
20 25 30

Thr Gln Asp Phe Glu Glu Ala Cys Arg Glu Leu Lys Leu Gly Glu Leu
35 40 45

Leu His Asp Lys Leu Phe Gly Leu Phe Glu Ala Met Ser Ala Ile Glu
50 55 60

Met Met Asp Pro Lys Met Asp Ala Gly Met Ile Gly Asn Gln Val Asn
65 70 75 80

Arg Lys Val Leu Asn Phe Glu Gln Ala Ile Lys Asp Gly Thr Ile Lys
85 90 95

Ile Lys Asp Leu Thr Leu Pro Glu Leu Ile Gly Ile Met Asp Thr Cys
100 105 110

Phe Cys Cys Leu Ile Thr Trp Leu Glu Gly His Ser Leu Ala Gln Thr
115 120 125

Val Phe Thr Cys Leu Tyr Ile His Asn Pro Asp Phe Ile Glu Asp Pro
130 135 140

Ala Met Lys Ala Phe Ala Leu Gly Ile Leu Lys Ile Cys Asp Ile Ala
145 150 155 160

Arg Glu Lys Val Asn Lys Ala Ala Val Phe Glu Glu Glu Asp Phe Gln

165 170 175

Ser Met Thr Tyr Gly Phe Lys Met Ala Asn Ser Val Thr Asp Leu Arg
180 185 190

Val Thr Gly Met Leu Lys Asp Val Glu Asp Asp Met Gln Arg Arg Val
195 200 205

Lys Ser Thr Arg Ser Arg Gln Gly Glu Glu Arg Asp Pro Glu Val Glu
210 215 220

Leu Glu His Gln Gln Cys Leu Ala Val Phe Ser Arg Val Lys Phe Thr
225 230 235 240

Arg Val Leu Leu Thr Val Leu Ile Ala Phe Thr Lys Lys Glu Thr Ser
245 250 255

Ala Val Ala Glu Ala Gln Lys Leu Met Val Gln Ala Ala Asp Leu Leu
260 265 270

Ser Ala Ile His Asn Ser Leu His His Gly Ile Gln Ala Gln Asn Asp
275 280 285

Thr Thr Lys Gly Asp His Pro Ile Met Met Gly Phe Glu Pro Leu Val
290 295 300

Asn Gln Arg Leu Leu Pro Pro Thr Phe Pro Arg Tyr Ala Lys Ile Ile
305 310 315 320

Lys Arg Glu Glu Met Val Asn Tyr Phe Ala Arg Leu Ile Asp Arg Ile
325 330 335

Lys Thr Val Cys Glu Val Val Asn Leu Thr Asn Leu His Cys Ile Leu
340 345 350

Asp Phe Phe Cys Glu Phe Ser Glu Gln Ser Pro Cys Val Leu Ser Arg
355 360 365

Ser Leu Leu Gln Thr Thr Phe Leu Val Asp Asn Lys Lys Val Phe Gly
370 375 380

Thr His Leu Met Gln Asp Met Val Lys Asp Ala Leu Arg Ser Phe Val
385 390 395 400

```
Asp Pro Pro Val Leu Ser Pro Lys Cys Tyr Leu Tyr Asn Asn His Gln
                405                 410                 415

Ala Lys Asp Cys Ile Asp Ser Phe Val Thr His Cys Val Arg Pro Phe
            420                 425                 430

Cys Ser Leu Ile Gln Ile His Gly His Asn Arg Ala Arg Gln Arg Asp
        435                 440                 445

Lys Leu Gly His Ile Leu Glu Glu Phe Ala Thr Leu Gln Asp Glu Phe
    450                 455                 460

Met Thr Phe Tyr Phe Asn Arg Ala Glu Lys Val Asp Ala Ala Leu His
465                 470                 475                 480

Thr Met Leu Leu Lys Gln Glu Pro Gln Arg Gln His Leu Ala Cys Leu
                485                 490                 495

Gly Thr Trp Val Leu Tyr His Asn Leu Arg Ile Met Ile Gln Tyr Leu
            500                 505                 510

Leu Ser Gly Phe Glu Leu Glu Leu Tyr Ser Met His Glu Tyr Tyr Tyr
        515                 520                 525

Ile Tyr Trp Tyr Leu Ser Glu Phe Leu Tyr Ala Trp Leu Met Ser Thr
    530                 535                 540

Leu Ser Arg Ala Asp Gly Ser Gln Met Ala Glu Glu Arg Ile Met Glu
545                 550                 555                 560

Glu Gln Gln Lys Gly Arg Ser Ser Lys Lys Thr Lys Lys Lys Lys Lys
                565                 570                 575

Val Arg Pro Leu Ser Arg Glu Ile Thr Met Ser Gln Ala Tyr Gln Asn
            580                 585                 590

Met Cys Ala Gly Met Phe Lys Thr Met Val Ala Phe Asp Met Asp Gly
        595                 600                 605

Lys Val Arg Lys Pro Lys Phe Glu Leu Asp Ser Glu Gln Val Arg Tyr
    610                 615                 620

Glu His Arg Phe Ala Pro Phe Asn Ser Val Met Thr Pro Pro Pro Val
625                 630                 635                 640
```

```
His Tyr Leu Gln Phe Lys Glu Met Ser Asp Leu Asn Lys Tyr Ser Pro
                645                 650                 655

Pro Pro Gln Ser Pro Glu Leu Tyr Val Ala Ala Ser Lys His Phe Gln
            660                 665                 670

Gln Ala Lys Met Ile Leu Glu Asn Ile Pro Asn Pro Asp His Glu Val
        675                 680                 685

Asn Arg Ile Leu Lys Val Ala Lys Pro Asn Phe Val Val Met Lys Leu
    690                 695                 700

Leu Ala Gly Gly His Lys Lys Glu Ser Lys Val Pro Pro Glu Phe Asp
705                 710                 715                 720

Phe Ser Ala His Lys Tyr Phe Pro Val Val Lys Leu Val
                725                 730
```

<210> 146
<211> 177
<212> PRT
<213> Homo sapien

<400> 146

```
Met Phe Phe Cys Val Gly Gly Tyr His Leu Val Phe Ser Arg Ser Ala
1               5                   10                  15

Phe Phe Val Arg Gly Arg Cys Gly Gly Phe Ser Arg Arg Leu Leu Ala
            20                  25                  30

Leu Ser Val Ala Gly Leu Gly Val Gly Leu Ser Gly Val Phe Met Val
        35                  40                  45

Asp Ala Gly Trp Phe Ile Arg Ser Ser Gly Leu Leu Leu Phe Phe Cys
    50                  55                  60

Leu Phe Ser Ser Arg Leu Phe Ser Pro Ser Cys Ser Leu Arg Pro Arg
65                  70                  75                  80

Ser Leu Leu Cys Ala Ala Val Ala Ser His Val Cys Pro Arg Arg Cys
                85                  90                  95

Val Phe Trp Ser Phe Ser Val Leu Ala Met Cys Leu Cys Val Cys Val
            100                 105                 110
```

Leu Leu Leu Leu Trp Ala Ala Pro Arg Val Val Val Thr Val Gly Ser
115 120 125

Leu Ser Pro Leu Cys Cys Cys Gly Ile Cys Glu Ala Gly Asn His Phe
130 135 140

Thr Pro Gly Asn His Ala Met Ser Pro Gly Tyr Pro Gln Leu Ile Gln
145 150 155 160

Thr Ser Lys Phe Trp Gly Gln Val Ile Leu Arg Pro Pro Arg Trp Phe
165 170 175

Phe

<210> 147
<211> 56
<212> PRT
<213> Homo sapien

<400> 147

Met Gln Asp Pro Val Leu Ser Asp Thr Arg Ser Ser Leu Gly Gly Val
1 5 10 15

Leu Gly Leu Leu Thr His Asn Phe Phe Thr Leu Val Leu Phe Trp Ser
20 25 30

Leu Ile Leu Ala Arg Asn Gln Pro Phe Gln Phe Leu Phe Lys Pro Lys
35 40 45

Lys Pro Leu Leu Val Gln Pro Gly
50 55

<210> 148
<211> 42
<212> PRT
<213> Homo sapien

<400> 148

Met Thr Asn Gly Arg Met Gly Leu Arg Cys Met Pro Ser Gly Ala Ser
1 5 10 15

Val Met Asp Ala Gly Arg Arg Ala Gly Thr Ala Asp Phe Gln Ser Lys
20 25 30

Asp Ile Tyr Leu Leu Tyr His Ile Ala Ser
35 40

<210> 149
<211> 27
<212> PRT
<213> Homo sapien

<400> 149

Met Cys Val Trp Cys Val Trp Tyr Val Val Tyr Val Val Cys Gly Val
1 5 10 15

Cys Arg Val Cys Gly Gly Tyr Thr Thr Leu Tyr
20 25

<210> 150
<211> 186
<212> PRT
<213> Homo sapien

<400> 150

Lys Ile Phe Leu Lys Gln Ile Lys Asp Ile Asn Lys Ala Lys Ser Ile
1 5 10 15

Tyr Leu Gln Cys Ile Tyr Leu Thr Lys Asp Ser Tyr Pro Glu Tyr Ile
20 25 30

Lys Ser Pro Tyr Lys Ser Met Thr Lys Asp Ile Ala Lys Thr Asn Lys
35 40 45

Thr Arg Cys Thr Met Ala Ser Gln His Ile Leu Lys Arg Phe Ser Ile
50 55 60

Ser Leu Val Ile Arg Glu Met Gln Lys Glu Thr Ile Met Arg Gly His
65 70 75 80

His Met Ile Thr Thr Leu Ala Lys Ile Lys Asn Thr Gln Asn Ala Lys
85 90 95

Cys Trp Ala Glu Cys Arg Glu Thr Gly Thr Arg Val His Cys Trp Trp
100 105 110

Glu Cys Lys Ile Val His Leu Leu Trp Lys Arg Val Trp Glu Phe Leu
115 120 125

Ala Lys Leu Asn Val Glu Leu Pro Tyr Asp Pro Ala Ile Pro Leu Leu

130 135 140

Cys Ile Asp Pro Arg Glu Leu Lys Thr Tyr Gly Gln Asn Thr Thr Cys
145 150 155 160

Ser Ala Met Phe Ile Met Thr Leu Phe Met Ile Ala Lys Lys Trp Lys
165 170 175

Gln Pro Lys Cys Pro Ser Arg Cys Pro Ser
180 185

<210> 151
<211> 201
<212> PRT
<213> Homo sapien

<400> 151

Met Pro Ser Pro Ser Arg Gly Val Ser Ile Leu Arg Ala Leu Pro Cys
1 5 10 15

Ser Leu Val Arg Val Arg Gly Cys Phe Val Arg Leu Gly Ser Leu Pro
20 25 30

Cys Pro Val Leu Val Arg Cys Tyr Phe Leu Phe Arg Leu Pro Phe Val
35 40 45

Leu Ser Ala Ala Pro Gly Leu Pro Arg Leu Ser Pro Pro Ala Leu Ser
50 55 60

Pro Pro Cys Pro Leu Arg Pro Ala Pro Ser Phe Leu Val Leu Leu Val
65 70 75 80

Val Asp Val Trp Gly Asn Cys Ala Glu Ala Arg Asn Asn Pro Gln Cys
85 90 95

Leu Ala Thr Thr Thr Ala Lys His Thr Pro Phe Val Thr Pro Met Glu
100 105 110

Val Tyr Leu Leu Leu Lys Ala Leu Leu Arg Ser Arg Lys Pro Phe Pro
115 120 125

Phe Pro Arg Gly Gly Pro Lys Leu Leu Gly Gly Pro Phe Pro Asn Gly
130 135 140

Pro Lys Arg Lys Thr Ala Val Ser Arg Val Thr Lys Arg Glu Leu Gly

145 150 155 160

Phe Thr Val Arg Val Gly His Asn His Val Trp Ala Cys Arg Gly Asn
165 170 175

Thr Ala Gln Lys Ser Gly Pro Pro His Thr Pro Lys Trp Glu Lys Pro
180 185 190

Gln Ala Arg Ala Leu Pro Asn Gly Leu
195 200

<210> 152
<211> 27
<212> PRT
<213> Homo sapien

<400> 152

Met Asp Ser Val Val Ala Thr Arg Tyr Phe Leu Gly Gly Pro Ser His
1 5 10 15

Pro Arg Glu Leu Cys Leu Pro Arg Thr Leu Lys
20 25

<210> 153
<211> 17
<212> PRT
<213> Homo sapien

<400> 153

Met Phe Asn Lys Val Glu Ser Thr Gly Gln Lys Lys Lys Lys Lys Lys
1 5 10 15

Lys

<210> 154
<211> 29
<212> PRT
<213> Homo sapien

<400> 154

Met Val Val Pro Gly Lys Leu Cys Lys Gly Leu Pro Tyr Lys Thr Ala
1 5 10 15

Ile Leu Thr Phe Cys Pro Thr Cys Thr Tyr Gly Ser Tyr
20 25

<210> 155
<211> 53
<212> PRT
<213> Homo sapien

<400> 155

Met Ile Val Leu Leu His Ser Ser Leu Gly Asp Thr Ala Ser Ser Cys
1 5 10 15

Phe Gln Thr Thr Thr Arg Lys Gln Asn Lys Lys Lys Lys Lys Lys Lys
20 25 30

Lys Lys Arg Leu Gly Tyr Trp Ala Ser Ser Gly Gly Gly Phe Phe Ser
35 40 45

Arg Pro Ser Pro Ile
50

<210> 156
<211> 81
<212> PRT
<213> Homo sapien

<400> 156

Trp Lys Gln Glu Leu Ala Val Ser Pro Arg Leu Glu Cys Ser Ser Thr
1 5 10 15

Ile Ile Ala His Ser Ser Leu Asp Leu Leu Cys Ala Asn Leu Pro Pro
20 25 30

Ala Ser Gly Ser Ala Val Ala Glu Thr Thr Gly Ala Cys Tyr His Thr
35 40 45

Trp Leu Ile Phe Lys Lys Met Phe Leu Glu Met Gly Ser His Asp Val
50 55 60

Ala Arg Ala Asp Leu Glu Leu Leu Ala Ser Asn Asn Tyr Ser Thr Ser
65 70 75 80

Ala

<210> 157
<211> 71
<212> PRT
<213> Homo sapien

<400> 157

Met His Ala Ser Cys Leu Lys Val Lys Asp Glu Gln Arg His His Trp
1 5 10 15

Thr Lys Leu Ser Trp Phe Ala Met Asn His Leu Ser Glu Gln Ala Asp
20 25 30

Asn Thr Pro Arg Tyr Ala Phe Ile Ser Thr Val Gly Thr Tyr Glu His
35 40 45

Gly Ile Pro Ile Ser Lys Ile Ser Asp Leu Phe Ser Leu Ser Val Arg
50 55 60

Thr Trp Tyr Val His Glu Gln
65 70

<210> 158
<211> 108
<212> PRT
<213> Homo sapien

<400> 158

Phe Tyr Leu Phe Met Lys Gln Gly Leu Thr Leu Ser Pro Arg Leu Glu
1 5 10 15

Cys Asn Gly Met Ile Leu Ala His Cys Ser Leu Arg Leu Leu Gly Ser
20 25 30

Ser Asp Ser Leu Ala Ser Ala Ser Ala Val Ala Gly Thr Thr Gly Thr
35 40 45

Arg His His Ala Gln Arg Asn Phe Phe Val Phe Leu Val Glu Met Gly
50 55 60

Ser His His Val Ala Thr Arg Leu Val Ser Asn Ile Val Thr Ser Glu
65 70 75 80

Ala Asp Pro Thr Cys Pro Ala Ala Ser Arg Arg Val Leu Gly Ile Thr
85 90 95

Ser Ala Thr Ser His Tyr Ala Trp Thr Ser Ile Val
100 105

<210> 159

<211> 279
<212> PRT
<213> Homo sapien

<400> 159

Met Leu Ala Ala Pro Phe Trp Leu Leu Phe Ser Asp Phe Gln Leu Ser
1 5 10 15

Phe Pro Ile Gln Pro His His Thr Thr Gln Ser Cys Lys Cys His Ser
20 25 30

Pro Pro Ser Leu Cys Leu Pro Pro His Pro Ser Pro Leu His Pro Ser
35 40 45

Ser Pro Ser His Pro Arg Pro Ala Arg His Leu Leu Pro Leu Arg His
50 55 60

Pro Ser Thr Pro Pro Ser Pro Thr Ser Leu Pro Ala Leu Pro Ser Leu
65 70 75 80

Ser Pro Leu Ser Ser Ile Pro His His Pro Pro Ser Thr Thr Ala Ala
85 90 95

Ile Gln Leu Pro Pro Thr Pro His His Leu Arg Pro Thr His Asn Tyr
100 105 110

Ser Pro Ile Arg Ser Ser His Ser Thr Pro Ser Pro His Asn Thr Pro
115 120 125

Arg Pro Thr Pro Thr Pro Pro Pro Pro Arg Ile His Tyr Thr Thr Ile
130 135 140

Ser Pro Leu Asn Thr Thr Ser Pro Pro Leu His Ser Thr Leu Ser Ser
145 150 155 160

Pro Pro Pro Leu His Gln Tyr Asn Pro Ser Gln Tyr Ser Tyr Thr Ile
165 170 175

Ile Gln Thr Ala Thr Thr His Pro Gln Leu Ser His Thr Pro Met Arg
180 185 190

Thr Asn Asn His His Ser Ile Leu Tyr Pro Pro Ser Leu Ser Pro Pro
195 200 205

Pro Pro Arg Thr Arg His Thr Pro Pro Pro His His Arg His His Leu

210 215 220

Leu Leu Tyr Leu Leu Pro Pro Tyr Thr Arg Pro Pro Thr Pro Leu Arg
225 230 235 240

Pro His Ser Ser Ser Thr Ile Tyr Thr Pro Pro Ala Tyr Ser Leu Pro
245 250 255

Ile Thr Pro Thr Ile Ser Ser Leu Ser Pro Gln Leu Pro Pro Ser His
260 265 270

Tyr His Leu Thr Thr Gln His
275

<210> 160
<211> 50
<212> PRT
<213> Homo sapien

<400> 160

Met Gln Thr Val Gly Phe Ala Gln Asp Phe His Asn Thr Gly Phe Asn
1 5 10 15

Tyr Pro Ile Arg Asp Ser Gln Leu Gly Arg Asp Thr Leu Phe Arg Asn
20 25 30

Pro Asn Phe Pro Phe Arg Asp Ile Trp Phe Tyr Thr Leu Arg Phe Tyr
35 40 45

Ser Arg
50

<210> 161
<211> 91
<212> PRT
<213> Homo sapien

<400> 161

Met Tyr Asn Ser Tyr Val Ser Trp Gly Pro His Arg Pro Ser Thr Ile
1 5 10 15

Val Pro Thr Phe Leu Phe Arg Asp Ser Ala Gln Pro Ser Phe Thr Thr
20 25 30

Thr Arg Ala Arg Thr Ile His Val Val Ile Ser Leu Ser Leu Ser Asn
35 40 45

Arg Gly Ser Thr Phe Ser Gln Lys Thr Phe Leu Ile Thr Arg Leu Thr
50 55 60

His Leu Ile Asn Lys Ala Ala Leu Phe Cys Arg Glu Arg Glu Leu Phe
65 70 75 80

Leu Ile Ala Thr Gln Gly Leu Phe Ser Arg Leu
85 90

<210> 162
<211> 109
<212> PRT
<213> Homo sapien

<400> 162

Met Phe Leu Asn Trp Arg Tyr Gln Tyr His Glu Asn Met Tyr Asn Asp
1 5 10 15

Leu Glu Ile Gln Tyr Leu Cys Met Asp Ile Cys Phe Val Lys Phe Val
20 25 30

Ser Gly Asp Phe Val Glu Arg Glu Arg Asn His Phe Pro His Thr Thr
35 40 45

Gly Asn Thr Ala Met Ala Thr Arg Gly Asn Arg His Gln Arg Leu Phe
50 55 60

Phe Phe Val Leu Tyr Met Phe Ser Ser Asp Gly Ser Leu Ala Val Leu
65 70 75 80

Pro Gly Trp Ser Ala Val Ala Arg Ser Arg Gly Ser Leu Gln Pro Leu
85 90 95

Thr Pro Gly Ser Thr Asp Ser Pro Gly Ser Ala Ser Gln
100 105

<210> 163
<211> 44
<212> PRT
<213> Homo sapien

<400> 163

Met Thr Met Gln Ala Thr Pro Thr Leu Ser Ser Pro Met Asn Thr Pro
1 5 10 15

Pro Gly Leu Arg Val Met Phe Trp Trp Trp Arg Ile Val Glu Ala Gly
20 25 30

Ile Ser Gln Cys Leu Thr His His Gly Lys His Gly
35 40

<210> 164
<211> 53
<212> PRT
<213> Homo sapien

<400> 164

Met Asn Thr Ala Asn Gln Pro Asn Glu Asn Ser Lys Arg Ser Pro Arg
1 5 10 15

Ser Glu Thr Asp Gly Gly Arg Pro Pro His Arg Arg Leu Ser Arg Lys
20 25 30

Gln Tyr Thr Arg Gln Leu Asp Pro Pro Trp Lys Arg Pro His His Glu
35 40 45

Ser Val Leu His Cys
50

<210> 165
<211> 60
<212> PRT
<213> Homo sapien

<400> 165

Met Asp Pro Leu His Cys Pro Phe Thr Thr Ala Ala Thr Ser Leu Ser
1 5 10 15

Tyr Thr Leu Thr Pro Thr Cys Gly Tyr His Cys Ser Val Leu His Leu
20 25 30

Cys Asn Phe Val Ile Ser Arg Met Leu Tyr Glu Trp Asn His Thr Glu
35 40 45

Cys Asn Leu Thr Arg Leu Ile Phe Phe His Ser Ala
50 55 60

<210> 166
<211> 213
<212> PRT
<213> Homo sapien

<400> 166

Ser Asn Arg Gly Ile Leu Ser Arg Ile Tyr Lys Lys Pro Leu Lys Thr
1 5 10 15

Gln Ala Ala Lys Glu Gln Met Thr Ala Ile Glu Asn Arg Gln Lys Thr
20 25 30

Ala Arg His Phe Thr Glu Glu Asp Thr Ala Met Ala Asn Ala His Thr
35 40 45

Lys Arg Tyr Ser Thr Ser Leu Ala Ile Glu Met Gln Ile Lys Thr Thr
50 55 60

Cys Gly Ile Ile Thr Thr Ser Met Ala Met Val Lys Ile Lys Asn Ser
65 70 75 80

Ser Asn Thr Lys Cys Trp Ala Gly Cys Glu Glu Thr Gly Ser Ile Ile
85 90 95

His Cys Cys Leu Asn Cys Met Ser Gly Cys Met Ala Lys Val Glu Pro
100 105 110

Leu Trp Lys Lys Ser Ala Gly Ser Phe Leu Gln Lys Tyr Met Cys Leu
115 120 125

Pro Tyr Asn Pro Thr Val Ala Leu Leu Ser Ile Tyr Pro Glu Asn Glu
130 135 140

Asn Val Cys Ser His Lys Thr Cys Thr Ala Met Phe Thr Ala Ala Phe
145 150 155 160

Ile Arg Ala Lys Asn Ala Lys Gln Leu Leu Cys Pro Leu Val Gly Glu
165 170 175

Trp Leu Ser Lys Leu Trp Tyr Ile His Thr Met Glu Tyr Tyr Ser Ala
180 185 190

Ile Lys Arg Asn Cys Pro His Phe Thr Thr Met Gln Tyr Met His Val
195 200 205

Arg Asn Leu Tyr Leu
210

<210> 167
<211> 127
<212> PRT
<213> Homo sapien

<400> 167

Met Ser Ile Gly Leu Asn Phe Thr Pro Arg Met Val Ala Arg Asp Met
1 5 10 15

Val Tyr Phe Val Pro Ile Leu Trp Thr Trp Arg Thr His Ala Ile Asp
20 25 30

Tyr Ala Lys Arg Arg Glu Thr Asn Thr Trp Val His Thr Pro Lys Ile
35 40 45

Pro Ala Leu Lys Arg Arg His Ser Ser Gly Thr Ile Ser Ala Thr Asn
50 55 60

Trp Gly Gly Leu Phe Thr Gln Gly Cys Lys Val Gly Lys Glu Lys Pro
65 70 75 80

Ser Leu Pro Leu Thr Ser His Glu Gln Phe Cys Ala Gly Val Tyr Pro
85 90 95

Ile Asn Thr Thr Gln Arg Thr Ile Ile Pro Pro Arg Gly Leu Leu Pro
100 105 110

Ser Leu Ser Pro Leu Pro Gly Glu Phe Thr Phe Phe Val Met Trp
115 120 125

<210> 168
<211> 60
<212> PRT
<213> Homo sapien

<400> 168

Met Asp Pro Leu His Cys Pro Phe Thr Thr Ala Ala Thr Ser Leu Ser
1 5 10 15

Tyr Thr Leu Thr Pro Thr Cys Gly Tyr His Cys Ser Val Leu His Leu
20 25 30

Cys Asn Phe Val Ile Ser Arg Met Leu Tyr Glu Trp Asn His Thr Glu
35 40 45

Cys Asn Leu Thr Arg Leu Ile Phe Phe His Ser Ala

50 55 60

<210> 169
<211> 211
<212> PRT
<213> Homo sapien

<400> 169

Pro Phe Ser Phe Leu Phe Arg Ala Leu Phe Ala Phe Phe Asp Pro Ala
1 5 10 15

Leu Ser Ile Leu Val Leu Ala Ile Ser Phe His Leu Pro Ile Asn Ser
20 25 30

Leu Ala Cys Leu Arg Glu Glu Ile His Lys Asp Leu Leu Val Thr Gly
35 40 45

Ala Tyr Glu Ile Ser Asp Gln Ser Gly Gly Ala Gly Gly Leu Arg Ser
50 55 60

His Leu Lys Ile Thr Asp Ser Ala Gly His Ile Leu Tyr Ser Lys Glu
65 70 75 80

Asp Ala Thr Lys Gly Lys Phe Ala Phe Thr Thr Glu Asp Tyr Asp Met
85 90 95

Phe Glu Val Cys Phe Glu Ser Lys Gly Thr Gly Arg Ile Pro Asp Gln
100 105 110

Leu Val Ile Leu Asp Met Lys His Gly Val Glu Ala Lys Asn Tyr Glu
115 120 125

Glu Ile Ala Lys Val Glu Lys Leu Lys Pro Leu Glu Val Glu Leu Arg
130 135 140

Arg Leu Glu Asp Leu Ser Glu Ser Ile Val Asn Asp Phe Ala Tyr Met
145 150 155 160

Lys Lys Arg Glu Glu Glu Met Arg Asp Thr Asn Glu Ser Thr Asn Thr
165 170 175

Arg Val Leu Tyr Phe Ser Ile Phe Ser Met Phe Cys Leu Ile Gly Leu
180 185 190

Ala Thr Trp Gln Val Phe Tyr Leu Arg Arg Phe Phe Lys Ala Lys Lys

195 200 205

Leu Ile Glu
210

<210> 170
<211> 49
<212> PRT
<213> Homo sapien

<400> 170

Met Val Ser Thr His Gln Arg Glu Thr Ser Tyr Asp His Gly Leu Thr
1 5 10 15

Pro Lys Leu Ser Gly Val Asn Leu Leu Lys Asn Lys Ile Arg Lys Thr
20 25 30

Glu Lys Cys Tyr Lys Pro Asn Asn Leu Lys Ile Gly Leu Lys Met Asn
35 40 45

Asn

<210> 171
<211> 146
<212> PRT
<213> Homo sapien

<400> 171

Met Phe Ala Val His Thr Ser Arg Phe Ala Val Gln Leu Arg Pro Phe
1 5 10 15

Val Leu Pro Leu Cys Phe Val Leu Thr His Phe Trp Leu Leu Thr Pro
20 25 30

Gly Pro Ile His Thr Lys Val Phe Pro Pro Thr Ser Asn Ile Arg Ala
35 40 45

Thr Arg Ser His Thr Thr Thr Thr Pro His Glu Pro Ala Leu His Thr
50 55 60

Pro His Pro Asp Pro Ala Pro Ser Thr Ser His Thr Pro His His Pro
65 70 75 80

Leu Asn Pro Pro Pro Thr His Thr Gln Pro Ser Leu Pro Thr Thr Pro
85 90 95

Leu Pro His Thr Pro His Thr Thr Thr Thr Pro His Thr Ser Thr Thr
100 105 110

Pro Thr Thr Pro Arg Thr Pro Thr His Pro Thr His Thr Pro Gln Pro
115 120 125

Thr Arg Pro His Thr His Pro His Thr Leu Thr Gln His Asn Asn Gln
130 135 140

Pro Pro
145

<210> 172
<211> 78
<212> PRT
<213> Homo sapien

<400> 172

Met Cys Thr Gln Ser Thr Thr Pro Gly Cys Asp Arg Thr Leu Gln Gly
1 5 10 15

Asp Thr Glu Ala His Trp Ser Arg Ala Arg Ala Pro Pro Lys Arg Thr
20 25 30

Ala Lys Gln Gly Ala Gln His Ser Thr Ala Pro Arg Gln Arg Ser Phe
35 40 45

Ser Arg Trp Pro Ser Ala Cys Pro Glu Gly His Ala Ala Gly Glu Arg
50 55 60

Gly Phe Gly Asn Pro Pro Ala Trp Thr Asp Thr Leu Arg Arg
65 70 75

<210> 173
<211> 78
<212> PRT
<213> Homo sapien

<400> 173

Met Tyr Lys Asn Glu Arg Tyr His Ala His His Thr Arg Val Val Gly
1 5 10 15

Glu Leu Pro Met Gly Leu Pro Ser Ser Arg Arg Arg Ser Ser Cys Arg
20 25 30

```
Thr Thr Cys Lys His Thr Ser Arg Glu Thr Leu Ser Gly Gln Thr Ser
        35                  40                  45

Ser Thr Thr Thr Ser Pro His Ala Arg Val Glu Leu Val Ile Ala Gln
    50                  55                  60

Ala Ser Gln Pro Val Cys Pro Ala Ile Ile Leu Leu Tyr Ile
65                  70                  75
```

<210> 174
<211> 111
<212> PRT
<213> Homo sapien

<400> 174

```
Met Leu Asp Thr Ile Glu Ser His Arg Gly Lys Ala Pro Ile Thr Lys
1               5                   10                  15

Arg Glu Arg Ser Ala Cys Phe Glu His Glu Leu Ser Lys Met Arg Glu
            20                  25                  30

Ser Met Arg Phe Lys Ala Ser Ala Ser Lys Leu Gly His Leu Val Asp
        35                  40                  45

Glu Lys Thr Tyr Gly His Pro Glu Gly Leu Trp Lys Thr Gln Pro Arg
    50                  55                  60

Thr His Ser Pro Gln Asp Thr Cys Leu Lys Ser Gly Ser Lys Pro Ser
65                  70                  75                  80

Cys Leu Gly Lys Glu Glu Gly Leu Gln Ser Ala Ala Asn Glu Arg Thr
                85                  90                  95

Leu Thr Lys Gly Lys Ile His Thr Arg Pro Asp Gln Pro Ile Arg
            100                 105                 110
```

<210> 175
<211> 134
<212> PRT
<213> Homo sapien

<400> 175

```
Met Cys Tyr Arg Glu Arg Cys Leu Leu Leu Val Glu Arg Thr His Thr
1               5                   10                  15
```

Leu Cys Ala Pro Thr Gln Cys Ser Val Val Gly Asp Asn Arg Ala Cys
20 25 30

Leu Ser Arg Leu Gln Arg Asp Ile Trp Ala Phe Phe Phe Phe Ser Arg
35 40 45

Arg Gly Ala Asp Thr Leu His Thr Arg Glu Val Cys Arg Ala Thr Tyr
50 55 60

Ile Ser Thr Gly Leu Ser Arg Glu Arg Tyr Leu Phe Ser Ser Leu Ser
65 70 75 80

Cys Gly Glu Asn Ser Leu Trp Cys Gly Asp His Thr Ala Arg His Lys
85 90 95

Arg Ser Ser Leu Ser Ser Val Lys His Ser Arg Arg Cys Leu His Lys
100 105 110

Asn Tyr Leu Ala Arg Pro Asn Arg Leu Leu Phe Phe Ile Phe Leu Asn
115 120 125

Ser Leu Trp Gly Gly Lys
130

<210> 176
<211> 234
<212> PRT
<213> Homo sapien

<400> 176

Met Phe Val Leu Leu Leu Cys Cys Leu Cys Leu Cys Leu Ser Val Cys
1 5 10 15

Phe Cys Leu Leu Ser Phe Gly Leu Cys Trp Val Leu Ser Cys Val Val
20 25 30

Leu Cys Val Val Phe Cys Phe Val Leu Phe Val Cys Val Leu Phe Phe
35 40 45

Val Leu Ser Leu Leu Phe Phe Leu Cys Cys Phe Cys Gly Phe Val Phe
50 55 60

Phe Leu Phe Cys Phe Val Cys Val Phe Phe Cys Cys Cys Val Leu Phe
65 70 75 80

```
Ser Phe Leu Leu Phe Val Phe Phe Ser Leu Cys Phe Phe Phe Val Leu
                85                  90                  95

Phe Ser Met Phe Leu Val Val Val Leu Phe Cys Leu Gly Leu Leu Phe
            100                 105                 110

Phe Phe Phe Cys Ser Val Ser Leu Cys Leu Phe Gly Phe Leu Leu Phe
        115                 120                 125

Phe Ser Phe Leu Phe Ser Leu Val Phe Val Val Leu Val Leu Phe Ala
    130                 135                 140

Cys Phe Trp Val Phe Ala Cys Cys Phe Cys Val Phe Phe Pro Phe Cys
145                 150                 155                 160

Leu Leu Val Phe Phe Phe Phe Leu Phe Phe Val Phe Arg Leu Phe Phe
                165                 170                 175

Phe Ser Phe Ser Leu Phe Ser Phe Phe Ala Phe Val Val Val Leu Cys
            180                 185                 190

Phe Leu Phe Phe Phe Leu Val Val Phe Phe Val Phe Phe Phe Phe Phe
        195                 200                 205

Phe Phe Ser Phe Ser Phe Phe Pro Leu Phe Phe Val Phe Phe Phe Phe
    210                 215                 220

Phe Phe Phe Phe Ser Phe Gly Ser Ser Arg
225                 230
```

<210> 177
<211> 123
<212> PRT
<213> Homo sapien

<400> 177

```
Met Ser Val Phe Ala Leu Ala Gly Arg Ser Cys Cys Cys Ser Val Cys
1               5                   10                  15

Cys Arg Val Ser Pro Val Cys Arg Leu Leu Cys Ser Cys Val Ser Phe
            20                  25                  30

Leu Cys Cys Leu Ala Ala Ser His Ile Ile Ser Ser Leu Gly Ile Arg
        35                  40                  45
```

Leu Leu Thr Val Tyr Leu Tyr Ser Cys Phe Ser Ile Phe Ala Cys Leu
50 55 60

Ala Phe Phe Phe Leu Ser Phe Phe Phe Val Gly Phe Leu Ile Phe Tyr
65 70 75 80

Glu Leu Gly Gly Thr His Cys Phe Pro Arg Arg Val Ile Phe Leu Leu
85 90 95

Pro Pro Val Leu Thr Pro His Arg Ser Phe Phe Phe Leu Phe Phe Val
100 105 110

Phe Phe Phe Ser Ser Val His Gln Thr Pro Leu
115 120

<210> 178
<211> 83
<212> PRT
<213> Homo sapien

<400> 178

Met Gly Arg Lys Thr Ile His Thr Gly Thr Leu Trp Pro Arg Leu Pro
1 5 10 15

Pro Thr Phe Phe Phe Phe Asp Ile Phe Phe Phe Ser Arg Arg Ser Leu
20 25 30

Ala Leu Leu Pro Arg Leu Glu Cys Ser Gly Ala Ile Ser Ala His Cys
35 40 45

Asn Phe Cys Leu His Lys Phe Lys Gln Phe Ser Cys Leu Ser Leu Gln
50 55 60

Ser Ser Trp Asp Tyr Arg Arg Val Pro Leu Cys Pro Ala Asn Phe Tyr
65 70 75 80

Ile Leu Met

<210> 179
<211> 71
<212> PRT
<213> Homo sapien

<400> 179

Met Arg Val Ser Thr Phe Val Arg Tyr Pro Arg Gly Asp Leu Thr Cys

1 5 10 15

Ala Gly Val Arg Ser Phe Ala Ser Arg Ser Leu Tyr His Val Val Arg
20 25 30

Leu Leu Val Gly Arg His Leu Ser Gly Asp Arg Val Ser Thr Pro Ser
35 40 45

Trp Pro Leu Ile Ala Ala Asp Cys Gln His Gly Leu Tyr Asp Leu Leu
50 55 60

Leu Ile Ser Ser Tyr Val Pro
65 70

<210> 180
<211> 84
<212> PRT
<213> Homo sapien

<400> 180

Met Phe Cys Leu Val Trp Gly Thr His His Leu Gly Cys Arg Arg Ala
1 5 10 15

Arg Gly Trp Leu Ile Thr Pro Pro Pro Cys Cys Ala Asn Thr Asn Pro
20 25 30

Arg Arg Gly Ile Thr Asn Ala Leu Ile Leu Glu Ala His Pro Trp Arg
35 40 45

Val Tyr Tyr Ala Pro Pro Thr Gly Phe Leu Gln Pro Arg Gly Gly His
50 55 60

Thr Ala Phe Asn Ser Val Val Ala Thr Arg Ser Cys Arg Gly Pro Pro
65 70 75 80

Thr Gly Gly Trp

<210> 181
<211> 74
<212> PRT
<213> Homo sapien

<400> 181

Met Glu Ser Thr Leu Arg Cys Ala Thr Pro Gly Pro Asp Thr Leu Gln
1 5 10 15

His Thr Gly Val Pro Gly Pro Ile Thr His Arg Glu Gln Val Gly Ser
20 25 30

Tyr Thr Thr Pro Leu Arg Ile Pro Pro Ala Ala Ala Asp Ser Gln Thr
35 40 45

Ala Val Tyr Asn Pro Leu Arg Arg Arg Arg Pro His Arg Ala Thr Pro
50 55 60

Arg Lys Pro Lys Thr Ile Thr Arg Lys Met
65 70

<210> 182
<211> 87
<212> PRT
<213> Homo sapien

<400> 182

Met Glu Leu Tyr His Arg Lys Glu Leu Glu Gly Leu Cys Tyr Cys Gly
1 5 10 15

Val Thr Phe Gly Leu Arg Ser Pro Gly Gln Ser Ala Arg Cys Cys Thr
20 25 30

Thr Arg Gly Asn His Cys Arg Cys His Pro Ala Pro Ala Pro Pro Pro
35 40 45

Gly Ala Pro Leu Arg Ile Ser Glu Lys Leu Lys Pro Ser Val Ser Leu
50 55 60

Gly Gly Phe Leu Arg Ser Ile Ile Ile Leu Leu Phe Asn Ser Ile Phe
65 70 75 80

Val Asn Ile Lys Ser Ser Phe
85

<210> 183
<211> 105
<212> PRT
<213> Homo sapien

<400> 183

Met Leu Lys Ser Phe Phe Phe Ser Leu Arg Gly Trp Gly Trp Arg Gly
1 5 10 15

```
Asp His Val Asn Phe Ser Gly Leu Gln Arg Lys Cys Gly Phe Val Asp
            20                  25                  30

Leu Gln Leu Phe Val Pro Phe Val Leu Ser Leu Cys Glu Ile Asn Thr
        35                  40                  45

Ser Lys Thr Phe Thr Pro Pro Leu Leu Ser Arg Gly Ala Tyr Ile Ser
    50                  55                  60

Arg Val Ala His Asn Ser Arg Val Ser Ala Gly Cys Glu Ser Val Phe
65                  70                  75                  80

Thr Arg Leu Pro Ile Pro Pro Lys Thr Ser Lys Lys Gly Val Pro Thr
                85                  90                  95

Lys Gly Thr Lys Glu Lys Lys Lys Pro
            100                 105
```

<210> 184
<211> 60
<212> PRT
<213> Homo sapien

<400> 184

```
Met Asp Pro Leu His Cys Pro Phe Thr Thr Ala Ala Thr Ser Leu Ser
1               5                   10                  15

Tyr Thr Leu Thr Pro Thr Cys Gly Tyr His Cys Ser Val Leu His Leu
            20                  25                  30

Cys Asn Phe Val Ile Ser Arg Met Leu Tyr Glu Trp Asn His Thr Glu
        35                  40                  45

Cys Asn Leu Thr Arg Leu Ile Phe Phe His Ser Ala
    50                  55                  60
```

<210> 185
<211> 218
<212> PRT
<213> Homo sapien

<400> 185

```
Ser Gly Leu Phe Gly Pro Pro Ala Arg Arg Gly Pro Phe Pro Leu Ala
1               5                   10                  15
```

Leu Leu Leu Phe Phe Leu Leu Gly Pro Arg Leu Val Leu Ala Ile Ser
20 25 30

Phe His Leu Pro Ile Asn Ser Arg Lys Cys Leu Arg Glu Glu Ile His
35 40 45

Lys Asp Leu Leu Val Thr Gly Ala Tyr Glu Ile Ser Asp Gln Ser Gly
50 55 60

Gly Ala Gly Gly Leu Arg Ser His Leu Lys Ile Thr Asp Ser Ala Gly
65 70 75 80

His Ile Leu Tyr Ser Lys Glu Asp Ala Thr Lys Gly Lys Phe Ala Phe
85 90 95

Thr Thr Glu Asp Tyr Asp Met Phe Glu Val Cys Phe Glu Ser Lys Gly
100 105 110

Thr Gly Arg Ile Pro Asp Gln Leu Val Ile Leu Asp Met Lys His Gly
115 120 125

Val Glu Ala Lys Asn Tyr Glu Glu Ile Ala Lys Val Glu Lys Leu Lys
130 135 140

Pro Leu Glu Val Glu Leu Arg Arg Leu Glu Asp Leu Ser Glu Ser Ile
145 150 155 160

Val Asn Asp Phe Ala Tyr Met Lys Lys Arg Glu Glu Glu Met Arg Asp
165 170 175

Thr Asn Glu Ser Thr Asn Thr Arg Val Leu Tyr Phe Ser Ile Phe Ser
180 185 190

Met Phe Cys Leu Ile Gly Leu Ala Thr Trp Gln Val Phe Tyr Leu Arg
195 200 205

Arg Phe Phe Lys Ala Lys Lys Leu Ile Glu
210 215

<210> 186
<211> 139
<212> PRT
<213> Homo sapien

<400> 186

Met Gln Val Val Ser Phe Leu Phe Pro Arg Ser Ser Cys Ser Asn Asp
1 5 10 15

Ser Ser Pro Gly Glu His His Gly Gly Asn Met His Ile Gly Arg Tyr
20 25 30

Gly Ser Ala Cys Ala Ile Val Arg Gly Ala Leu Trp Glu Asp Phe Ile
35 40 45

Met His Leu Ser Phe Arg Met Cys Pro Arg Val Ile Ser Glu Lys Glu
50 55 60

Gly Thr Val Glu Arg Ala Phe Leu Lys Gly Ile Lys Val Ala Leu Leu
65 70 75 80

Ile Ser Val Cys Arg Phe Met Ser Pro Ser Trp Ile Pro Trp Trp Ala
85 90 95

Pro Asn Asn Ala Ala Pro Lys Ile Gln Val Phe Arg Ile Ile Tyr Pro
100 105 110

Leu Leu Pro Tyr His Thr Gly Gly Thr Gly Thr Ser Gln Val Val Gly
115 120 125

Ser Arg Met Glu Val Gly Val Tyr Gly Val Arg
130 135

<210> 187
<211> 118
<212> PRT
<213> Homo sapien

<400> 187

Met Leu Trp Gly Trp Gly Pro Arg Val Ala Leu Gln Arg Leu Val Tyr
1 5 10 15

Ser Pro Ala Ser Leu Gly Gly Ala Arg Val Gly Val Val Ile His Gly
20 25 30

Trp Ser Asn Glu Tyr Leu Thr Thr Tyr Pro Ala Val Leu Thr Pro Phe
35 40 45

Glu Pro Arg Val Leu Tyr Leu Lys Lys Tyr Ser Pro Lys Gln Thr Gln
50 55 60

Ile Phe Ala Ala Val Gly Gly Gly Ala Pro Phe Gly Leu Ser Pro Arg
65 70 75 80

Tyr Pro Gly Gly Cys Gly Gly Thr Glu Lys Trp Gly Val Cys Pro Trp
85 90 95

Gly Gly Ala Ala Leu Leu Val Lys Pro Glu Lys Ser Ala Ser Leu Trp
100 105 110

Ala Pro Arg Val Asp Val
115

<210> 188
<211> 202
<212> PRT
<213> Homo sapien

<400> 188

Met Trp His Thr Ser Val Gly Thr Ser Leu His Leu Ser His Thr Glu
1 5 10 15

Phe Ser Arg Cys Gly Lys Arg Gly Met Ser Pro Thr Arg Cys Ala Leu
20 25 30

Trp Val Ala His Lys Asn Thr Gln Arg Arg Glu Glu Arg Val Trp Cys
35 40 45

Gly Val Val Asp Glu Gly Pro Val Gly Glu Arg Glu Arg His Thr Pro
50 55 60

Pro Cys Arg Glu Arg Ala Gly Glu Thr His Arg Trp Ser Ser His Thr
65 70 75 80

Cys Glu Thr Leu Ser Pro Thr Gly Gly Arg Glu Lys Cys Val Ala Pro
85 90 95

Gly Ser Pro Cys Ala His Thr Ile Lys Glu Gly Asp Asp Thr Gln Lys
100 105 110

Thr Met Cys Ala Arg Val Arg Lys Thr Ile Val Arg Glu Arg Gly Val
115 120 125

Val Gly Ala Ser Gly Arg Ala Arg Gly Gly Arg Leu Thr Arg Ala Pro
130 135 140

Val Arg Asn Leu Pro Glu Thr Thr Cys Val Trp Arg Gly Ala His Arg
145 150 155 160

Gly Arg Arg Gly Asp Ser His Arg Glu Trp Val Tyr Lys Glu Arg Cys
165 170 175

Val Arg His Thr Gln Leu Ala Cys Ala Arg Asn Thr His Ala Arg Arg
180 185 190

Lys Tyr Pro Arg Gly Ser Leu Ser Thr Gln
195 200

<210> 189
<211> 102
<212> PRT
<213> Homo sapien

<400> 189

Met Thr Ile Ser Ile Gly Leu Cys Asp Val Tyr Asn Gln Trp Thr Ser
1 5 10 15

Leu Arg Leu Gly Phe Pro Val Ile Gly Cys Lys Gln Tyr Ala Cys Ser
20 25 30

Ser Gly Phe Thr Asp Met Tyr Pro Cys Ser Thr Tyr Ile Ser Gly Arg
35 40 45

Pro Ala Asn Lys Pro Ser Gly Asn Gly Trp Arg Arg Arg Val Ala Tyr
50 55 60

Gly Arg Arg Arg Pro Gly Asp Ser Ser Arg Glu Asn Glu Pro Ala Ile
65 70 75 80

Thr Thr Val Gly Ile Val His Ser Lys Arg Asn Lys Pro Arg Trp Arg
85 90 95

Glu Leu Arg Ile Pro Ala
100

<210> 190
<211> 65
<212> PRT
<213> Homo sapien

<400> 190

Met Leu Leu Ser Ser Ser Arg Pro His Lys Asp Val Asp Ser Gln Asn

1 5 10 15

Ser Asp Pro Val Pro Ala Asp Asp Asp Ala Ala Arg Leu Gln Val Ile
20 25 30

Ser Tyr Thr Ile Val Gly Asp Gly Val Arg Leu Leu Glu Ala Ser Met
35 40 45

Phe Lys Glu Tyr Ile Arg Gln Leu His Ala Thr His Trp Ile Arg Ser
50 55 60

Pro
65

<210> 191
<211> 145
<212> PRT
<213> Homo sapien

<400> 191

Met Thr Val Val Tyr Ala Gln Thr Asn Lys Lys Lys Thr Lys Lys Thr
1 5 10 15

Lys Glu Thr Pro Trp Gly Val Thr Pro Tyr Gly Gly Pro Met Arg Arg
20 25 30

Cys Val Ser Pro Trp Val Val Glu Thr Val Cys Val Leu Ser Gly Asn
35 40 45

Thr Asn Ile Leu Pro Pro His Asn Ile Leu Arg Arg Pro Gln Thr Gln
50 55 60

Lys His Thr Thr His Asn Pro Arg Thr Thr Leu Gln Gln Thr Thr Pro
65 70 75 80

Glu Lys Glu Leu Val Ala Ala Gln Val Lys Gln Gly Ala Pro Ala Ser
85 90 95

Pro Gln Lys Thr Pro Ile Glu Gln Cys Arg Lys Lys Arg Ser Thr Gly
100 105 110

Arg Glu Arg Leu Met Pro Gln Leu Glu His Glu Glu Lys Pro Asn Cys
115 120 125

Asn Leu Pro Thr Lys Cys Asp Glu Ile Arg Gln Glu Ala Ser Arg Arg

130 135 140

Ala
145

<210> 192
<211> 167
<212> PRT
<213> Homo sapien

<400> 192

Met Val Pro Phe Gly Val Phe Val Leu Cys Ser Arg Val Leu Phe Ser
1 5 10 15

Leu Val Leu Val Ala Phe Cys Phe Cys Leu Leu Leu Phe Phe Ser Ser
20 25 30

Phe Phe Ser Leu Val Arg Ser Phe Ser Phe Val Phe Phe Cys Cys Cys
35 40 45

Phe Leu Ser Tyr Phe Pro Leu Leu Phe Cys Phe Phe Phe Leu Ile Leu
50 55 60

Leu Phe Leu Phe Leu Leu Cys Leu Val Leu Phe Pro Cys Leu Ser Ser
65 70 75 80

Tyr Phe Leu Ser Val Trp Phe Cys Phe Val Val Leu Phe Ser Val Ala
85 90 95

Tyr Val Ser Cys Leu Ser Phe Ser Ser Phe Phe Ala Phe Phe Pro His
100 105 110

Leu Phe Phe Phe Phe Leu Ser Phe Leu Cys Phe Pro Leu Leu Leu Leu
115 120 125

Ser Leu Val Ser Ser Phe Val Trp Phe Leu Ser Leu Ser Pro Pro Cys
130 135 140

Leu Phe Phe Ser Ser Ser Phe Phe Val Ser Leu Ser Phe Val Phe His
145 150 155 160

Ser Pro Pro Ala Cys Leu Arg
165

<210> 193

<211> 151
<212> PRT
<213> Homo sapien

<400> 193

Met Trp Phe Cys Ile Phe Pro Leu Leu Ala Cys Leu Pro Ser Leu Ala
1 5 10 15

Phe Leu Phe Ser Phe Ala Ser Arg Leu Cys Leu Ser Val Pro Cys Val
20 25 30

Phe Ala Ser Thr Asp Leu Leu Pro Gly Phe Ser Trp Leu Ala Tyr Ser
35 40 45

Pro Val Asp Cys Leu Phe Ala Trp Glu Leu Phe Arg Leu Leu Leu Ser
50 55 60

Pro Leu Val Ser Val Val Gly Ser Trp Phe Leu Ala Leu Cys Ser Leu
65 70 75 80

Ala Cys Val Arg Leu Val Ser Ser Phe Glu Ser His Ala Gly Val Trp
85 90 95

Trp Cys Val Cys Val Val Val Ala Leu Gln Tyr Cys Leu Ser Leu Val
100 105 110

Leu Leu Ser Leu Ser Phe Val Ser Asp Val Leu Ser Tyr Phe Ser Leu
115 120 125

Gly Leu Leu Gln Cys Phe Ser Val Leu Gly Leu Ser Val Leu Leu Met
130 135 140

Ser Leu Ile Ala Phe Tyr Leu
145 150

<210> 194
<211> 122
<212> PRT
<213> Homo sapien

<400> 194

Met Thr Leu Ser Glu Ile Ala Arg Gln Arg Thr Glu Pro Gln Lys Tyr
1 5 10 15

Asp Gln Lys Arg Glu Asn Lys Asn Pro Gln Arg Gln Thr Asp Lys Glu
20 25 30

Arg Thr Lys Met Asn Lys Lys Thr Lys Lys Lys Lys Asn Thr Arg Arg
35 40 45

Glu Arg Lys Lys Glu Thr Thr Arg Lys Thr Arg Asn Lys Glu Arg Ser
50 55 60

Glu Thr Asn Arg Thr Lys Glu Gln Gln Lys Gln Asn Glu Gln Lys Asn
65 70 75 80

Asn Gly Thr Thr Thr Pro Pro Arg Lys Pro Lys Gln Arg Lys Gln Lys
85 90 95

Arg Ala Pro Leu Ser Arg His Thr Asn Arg Glu Arg Lys Thr Lys Asp
100 105 110

Thr Asn Asn Gln Asn Thr His Ile Val Gly
115 120

<210> 195
<211> 90
<212> PRT
<213> Homo sapien

<400> 195

Met Cys Phe Phe Phe Cys Phe Val Phe Phe Leu Leu Leu Phe Phe Ala
1 5 10 15

Cys Val Cys Cys Val Phe Cys Met Phe Leu Phe Val Cys Val Leu Leu
20 25 30

Ala Gly Arg Ser Phe Phe Val Phe Met Phe Gly Ser Pro Leu Phe Ser
35 40 45

Leu Cys Val Ser Pro Ala Tyr Met Cys Val Cys Val Trp Arg Asp Met
50 55 60

Cys Glu Ser Ala Arg Tyr Ile Thr His Phe Tyr Thr His Thr Gly Glu
65 70 75 80

Thr His Ser Ile Cys Glu Thr Thr Gly Glu
85 90

<210> 196
<211> 310

<212> PRT
<213> Homo sapien

<400> 196

Met Thr Ala Thr Thr Ala Ser Cys Gly Gly Gly Asn Asn Thr Pro Ala
1 5 10 15

Val Pro Pro Thr Pro Arg Gly Glu Ala His Ile Ser Thr Leu Val Trp
20 25 30

Cys Phe Arg Asp Ile Pro Pro Ala Ala Glu Leu Leu Trp Ala Pro Leu
35 40 45

Gly Val Leu Tyr Phe Ile His Leu Phe Leu Pro Leu Cys Leu Trp Gly
50 55 60

Asp Pro Pro Ala Tyr Lys Val Ile Ser Val Met Ile Leu His His Ile
65 70 75 80

Ile Val Phe Phe Leu Gly Glu Asp Thr Leu Gly Gly Asp Thr Thr Ser
85 90 95

Arg Gly Val Tyr Ala Pro Leu Pro His Met Arg Gly Ala Tyr Ser Ala
100 105 110

Pro Ser Glu Gly Ala His Pro Pro His Thr Leu Trp Ser His Ser Leu
115 120 125

Leu Cys Val Leu Pro Pro Ser Leu Ser Leu Ser Glu Arg Glu Ser Leu
130 135 140

Ser Thr Gln Pro His Thr His Arg Gly Ala His Thr His Ser Val Val
145 150 155 160

Cys Val Cys Leu Trp Ser Leu His Ser Gly Arg Leu Leu Tyr His Pro
165 170 175

Arg Gly Glu Thr Leu Cys Asp Asp Thr Ala Gly Ala Ala Leu Leu Glu
180 185 190

Arg Ala Thr Gln Ser Val Arg His Asn Ser Leu Thr Leu Phe Asn Arg
195 200 205

Asp Ala Arg Arg Val Trp Arg Asp Ala Thr Pro His Thr Arg Ser Leu
210 215 220

```
Ala His Thr His Arg Glu Arg His Thr His Thr His Val Asn Ala Ala
225                 230                 235                 240

Ala Thr Ala Thr Ala Leu Thr His Ser Arg Val Thr Arg Asp Ala Arg
                245                 250                 255

Ala Ala Ala Thr Ala Gly Arg Ser Val Ser Pro Thr Gln Arg Glu Ala
            260                 265                 270

Thr His Ser Ala Arg Ala His Ala Cys His His Ala His Ser Arg Glu
        275                 280                 285

Gly Glu Arg Asn Pro Leu Gly Glu Arg Arg His Thr Val Gly Ala Leu
    290                 295                 300

Thr Thr Arg Ser Val Thr
305                 310
```

<210> 197
<211> 122
<212> PRT
<213> Homo sapien

<400> 197

```
Met Phe Lys Ser Leu Asn Gln Tyr Arg Thr Leu Thr Pro Ser Gly Asn
1               5                   10                  15

Ser Asp Leu Pro Ser Ala Lys Leu Ser Arg Gln Ile Arg Phe Thr Ala
            20                  25                  30

Lys Thr Pro Pro Phe Thr Gln Tyr Thr Thr Arg Pro His Thr Leu Tyr
        35                  40                  45

Leu Ser Val Pro Cys Thr Leu Ser Ser Arg Ser Ser Asp Phe Arg His
    50                  55                  60

Thr Leu Glu Val Gly Lys Leu Leu Leu Met Leu Pro Leu Thr Gln Ser
65                  70                  75                  80

Ile Arg Phe Asp Arg Tyr Ser Cys Met Gln Leu Gln Lys Val Ser Tyr
                85                  90                  95

Phe Ser Ser Asp Ala Met Ser Thr Ala Ala Asp Gln Arg Tyr His Gly
            100                 105                 110
```

Val Tyr Arg Ile Cys Val Tyr Leu Lys Arg
115 120

<210> 198
<211> 91
<212> PRT
<213> Homo sapien

<400> 198

Met Glu Ser Arg Ser Val Ala Gln Ala Gly Val Gln Trp Arg Asp Leu
1 5 10 15

Ser Ser Leu Gln Leu Leu Pro Pro Gly Ile Lys Arg Phe Ser Cys Leu
20 25 30

Ser Leu Leu Ser Ser Trp Asp Tyr Arg His Pro Pro Pro Cys Pro Ala
35 40 45

Asn Phe Cys Val Phe Ser Arg Asp Gly Leu Ser Pro Cys Trp Pro Val
50 55 60

Trp Pro Arg Thr Pro Asp Pro Arg Ile Leu Leu Pro Gln Pro Pro Lys
65 70 75 80

Val Leu Gly Leu Gln Thr Cys Pro Gly Gly Arg
85 90

<210> 199
<211> 107
<212> PRT
<213> Homo sapien

<400> 199

Met Thr Lys Gln Ser Ser Ile Thr Pro Pro Lys Asp His Val Ser Ser
1 5 10 15

Pro Ala Met Asp Pro Asn Gln Glu Glu Ile Ser Glu Leu Pro Glu Lys
20 25 30

Glu Phe Arg Arg Pro Ile Ile Gln Leu Leu Lys Glu Thr Pro Asp Lys
35 40 45

Gly Val Asn Gln Leu Lys Gly Ile Lys Ile Ile Ile Gln Asp Met Asp
50 55 60

Glu Lys Val Ser Arg Glu Ile Asp Ile Ile Asn Lys Asn Gln Ser Gln
65 70 75 80

Leu Leu Glu Val Lys Asp Ile Leu Arg Glu Ile Gln Asn Thr Leu Ala
85 90 95

Ser Phe Asn Asn Gly Leu Glu Gln Val Glu Glu
100 105

<210> 200
<211> 32
<212> PRT
<213> Homo sapien

<400> 200

Met Leu Val Cys Lys Val Leu Leu Arg Arg Ile Gln Asn Thr Lys Leu
1 5 10 15

Leu Phe Phe Thr Cys Phe Phe Lys Phe Thr Tyr Leu Tyr Leu His Leu
20 25 30

<210> 201
<211> 342
<212> PRT
<213> Homo sapien

<400> 201

Leu Leu Lys Leu Leu Gln Val Leu Ile Val Leu Glu His His Leu Gly
1 5 10 15

Arg Ala His Glu Glu Ala Glu Asn Gln Pro Asp Leu Ser Arg Glu Trp
20 25 30

Gln Arg Ala Leu Asn Phe Gln Gln Ala Ile Ser Ala Leu Gln Tyr Val
35 40 45

Gln Pro His Pro Leu Thr Ser Gln Gly Leu Leu Val Ser Ala Val Val
50 55 60

Arg Gly Leu Gln Pro Ala Tyr Gly Tyr Gly Met His Pro Ala Trp Val
65 70 75 80

Ser Leu Val Thr His Ser Leu Pro Tyr Phe Gly Lys Ser Leu Gly Trp
85 90 95

```
Thr Val Thr Pro Phe Val Val Gln Ile Cys Lys Asn Leu Asp Asp Leu
            100                 105                 110

Val Lys Gln Tyr Glu Ser Glu Ser Val Lys Leu Ser Val Ser Thr Thr
        115                 120                 125

Ser Lys Arg Glu Asn Ile Ser Pro Asp Tyr Pro Leu Thr Leu Leu Glu
    130                 135                 140

Gly Leu Thr Thr Ile Ser His Phe Cys Leu Leu Glu Gln Ala Asn Gln
145                 150                 155                 160

Asn Lys Lys Thr Met Ala Ala Gly Asp Pro Ala Asn Leu Arg Asn Ala
                165                 170                 175

Arg Asn Ala Ile Leu Glu Glu Leu Pro Arg Thr Val Asn Thr Met Ala
            180                 185                 190

Leu Leu Trp Asn Val Leu Arg Lys Glu Glu Thr Gln Lys Arg Pro Val
        195                 200                 205

Asp Leu Leu Gly Ala Thr Lys Gly Ser Ser Ser Val Tyr Phe Lys Thr
    210                 215                 220

Thr Lys Thr Ile Arg Gln Lys Ile Leu Asp Phe Leu Asn Pro Leu Thr
225                 230                 235                 240

Ala His Leu Gly Val Gln Leu Thr Ala Ala Val Ala Ala Val Trp Ser
                245                 250                 255

Arg Lys Lys Ala Gln Arg His Ser Lys Met Lys Ile Ile Pro Thr Ala
            260                 265                 270

Ser Ala Ser Gln Leu Thr Leu Val Asp Leu Val Cys Ala Leu Ser Thr
        275                 280                 285

Leu Gln Thr Asp Thr Leu Leu His Leu Val Lys Glu Val Val Lys Arg
    290                 295                 300

Pro Pro Gln Val Lys Gly Gly Asp Glu Lys Ser Pro Leu Val Asp Ile
305                 310                 315                 320

Pro Val Leu Gln Phe Cys Tyr Ala Phe Leu Gln Arg Ala Tyr Ser Pro
                325                 330                 335
```

Pro Ser Ser Lys Asn Phe
340

<210> 202
<211> 221
<212> PRT
<213> Homo sapien

<400> 202

Gly Ser Trp Ala Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Gly
1 5 10 15

Ala Pro Gly Gln Arg Val Thr Ile Ser Cys Thr Gly Ser Ser Ser Asn
20 25 30

Ile Gly Ala Gly Tyr Asp Tyr Val His Trp Tyr Gln Gln Leu Pro Gly
35 40 45

Thr Ala Pro Lys Leu Met Ile Tyr Glu Val Ala Lys Arg Pro Ser Gly
50 55 60

Val Ser Asp Arg Phe Ser Gly Ser Lys Ser Gly Asn Thr Ala Ser Leu
65 70 75 80

Thr Ile Ser Gly Leu Gln Ala Glu Asp Glu Ala Asp Tyr Tyr Cys Cys
85 90 95

Ser Tyr Ala Gly Ser Tyr Thr Trp Val Phe Gly Gly Gly Thr Lys Leu
100 105 110

Thr Val Leu Gly Gln Pro Lys Ala Ala Pro Ser Val Thr Leu Phe Pro
115 120 125

Pro Ser Ser Glu Glu Leu Gln Ala Asn Lys Ala Thr Leu Val Cys Leu
130 135 140

Ile Ser Asp Phe Tyr Pro Gly Ala Val Thr Val Ala Trp Lys Ala Asp
145 150 155 160

Ser Ser Pro Val Lys Ala Gly Val Glu Thr Thr Thr Pro Ser Lys Gln
165 170 175

Ser Asn Asn Lys Tyr Ala Ala Ser Ser Tyr Leu Ser Leu Thr Pro Glu
180 185 190

```
Gln Trp Lys Ser His Lys Ser Tyr Ser Cys Gln Val Thr His Glu Gly
        195                 200                 205

Ser Thr Val Glu Lys Thr Val Ala Pro Thr Glu Cys Ser
    210                 215                 220
```

<210> 203
<211> 150
<212> PRT
<213> Homo sapien

<400> 203

```
Met Thr Val Arg Val Thr Tyr Thr Asn Val Leu Ser Glu Val Arg Arg
1               5                   10                  15

Pro Ile Pro Lys Tyr Ala Pro Met Cys Leu Val Leu His Ser Ile Leu
            20                  25                  30

Pro Tyr Pro Met His Ala Lys Cys Met Val Ser Thr Trp Cys Pro Asn
        35                  40                  45

Val Ser Ala Tyr Tyr Thr Lys Thr Thr Thr Cys Ser Thr His Asn Arg
    50                  55                  60

Cys Asn Met Gln Ser Thr Lys Gln Gly His Thr Ala Gln Leu Ala Ile
65                  70                  75                  80

Leu Thr Ile Glu Gln Ile Gln Ser Pro Asp Tyr Asn Met Leu Leu Thr
                85                  90                  95

His Gly Leu Leu Gln Ala Ala Gln Trp Asn Leu Gly Leu Ser Leu Lys
            100                 105                 110

Gln Gln Arg Tyr Ala Gln Leu Ala Ser Arg Thr Arg His Ala Asn Gly
        115                 120                 125

Ile Pro Ala Thr Gly Ala Arg Ser Ser Asn Asn His Glu His Arg Pro
    130                 135                 140

Glu Arg Arg Ala Leu Arg
145                 150
```

<210> 204
<211> 47
<212> PRT

<213> Homo sapien

<400> 204

Met Ser Val Ser Ile Ser Leu Val Ser Ser Pro Arg Gly Ser Thr Ala
1 5 10 15

Tyr His Pro Arg Ser Val Glu Ala Pro Lys Gly Leu Pro Phe Leu Ala
20 25 30

Val Arg Pro Cys Ala Asn Pro Cys Gln Asp Thr Pro Arg Gly Leu
35 40 45

<210> 205
<211> 130
<212> PRT
<213> Homo sapien

<400> 205

Met Arg His Arg Lys Arg Lys Ser Thr Arg Arg Lys Lys Arg Arg Arg
1 5 10 15

Ile Glu Glu Arg His Val Thr Glu Asn Arg Asp Gln Glu Arg Ser Lys
20 25 30

Asp Arg Pro Gln Arg Gln Asp Gly Gly Gly Glu Arg Lys Arg Ser Gln
35 40 45

Lys Lys Thr Lys Asn Glu Arg Ile Thr Glu Ile Asn Thr Ala Thr Arg
50 55 60

Glu Gln Thr Arg Gln Glu Gln Lys Lys His Lys Gln Gln Arg Glu Ala
65 70 75 80

Lys Arg Lys Lys Arg Lys Gly Arg Gln Gln Thr Lys Glu Thr Lys Arg
85 90 95

His Arg Gln Met Glu Arg Lys Arg Glu Gln His Arg Glu Glu Gly Arg
100 105 110

Lys Glu Ile Glu Thr Arg Ala Lys Arg Ala Arg Asn Lys Lys Arg Glu
115 120 125

Ala Arg
130

<210> 206
<211> 58
<212> PRT
<213> Homo sapien

<400> 206

Met Asn Asn Gly Arg Cys Val Asn Trp Ser Asn Thr Leu Leu His Trp
1 5 10 15

Thr Gln Trp Thr Pro Arg Cys Ala Lys His His Lys Lys Asp Gly Gly
20 25 30

Gln Arg Ser Thr Asp Gly His His Thr Thr Arg Ser Ile Thr Ser Glu
35 40 45

Asn Tyr Pro Arg Thr Asn Lys Glu Leu Lys
50 55

<210> 207
<211> 60
<212> PRT
<213> Homo sapien

<400> 207

Met Arg Leu Arg Cys Tyr Ile Cys Thr Leu Phe Phe Phe Phe Cys Phe
1 5 10 15

Phe Phe Phe Leu Ser Ser Arg Phe Val Ser Gly Met Cys Cys Trp Gly
20 25 30

Glu Leu Val Gly Ala Glu Ile Ser Thr Leu Val Thr His Arg Gly Asn
35 40 45

Thr Arg Leu Met Gly Pro Trp Leu Ser Pro Thr Arg
50 55 60

<210> 208
<211> 188
<212> PRT
<213> Homo sapien

<400> 208

Met Gln Asn Thr Thr Gly Val Thr Thr Gln Lys Arg Leu Glu Leu Gln
1 5 10 15

Ala Leu Tyr Thr Asn Cys Asp Gln Glu His Leu Leu Leu Thr Thr Ile
20 25 30

```
Ser Ser Ala Arg Arg His Lys Asn Met Val Cys Thr Arg Gly Val Asp
        35                  40                  45

Asn His His Leu Cys Ala Gly Leu Arg Gly Arg Arg Ala Thr His Ser
    50                  55                  60

Leu Ala Tyr Asn Ser Arg Cys Arg Thr Trp Arg Val Gly Leu Glu Thr
65                  70                  75                  80

Leu Arg Gly Cys Asn Thr Asp Val His Gly Ala Ser Gly Lys Gln Thr
                85                  90                  95

Arg Thr Gln Gln Arg Gly Glu Lys His Cys Phe Val Asn Arg Glu Asn
            100                 105                 110

Thr Arg Met Ile Lys Asn Arg Pro Thr Gly Ala Gly Gly Thr Ile Thr
        115                 120                 125

Thr Thr Glu Thr Leu Thr His Leu Gln Gly Gly Val Glu Gly Pro Leu
    130                 135                 140

Asp Thr Pro Leu Lys Pro Arg Lys Ser Asn Asn Asp Ala Thr Lys Pro
145                 150                 155                 160

Lys Ile Ala Thr His Ala Val Gln Ala Trp Ala Asp Thr Ala Arg Ser
                165                 170                 175

Gly Ser Pro Lys Lys Glu Lys His Pro Lys Lys Gln
            180                 185
```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2002 (22.08.2002)

PCT

(10) International Publication Number
WO 02/064788 A3

(51) International Patent Classification[7]: C12N 15/12, C07K 14/47, C12Q 1/68, C12N 5/10, C07K 16/18, G01N 33/53

(21) International Application Number: PCT/US01/45080

(22) International Filing Date:
20 November 2001 (20.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/252,054 20 November 2000 (20.11.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/252,054 (CIP)
Filed on 20 November 2000 (20.11.2000)

(71) Applicant *(for all designated States except US)*: **DIADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **MACINA, Roberto, A.** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **RECIPON, Herve** [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). **CHEN, Sei-Yu** [—/US]; 160 Mira Street, Foster City, CA 94404 (US). **SUN, Yongming** [CN/US]; 551 Shoal Drive, Redwood City, CA 94065 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US).

(74) Agents: **LICATA, Jane, Massey** et al.; Licata & Tyrrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
11 December 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/064788 A3

(54) Title: COMPOSITIONS AND METHODS RELATING TO LUNG SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic lung cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating lung cancer and non-cancerous disease states in lung, identifying lung tissue, monitoring and identifying and/or designing agonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered lung tissue for treatment and research.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/45080

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/47 C12Q1/68 C12N5/10 C07K16/18
G01N33/53

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, BIOSIS, CHEM ABS Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 00 08206 A (MACINA ROBERTO A ;SUN YONGMING (CN); YANG FEI (CN); RECIPON HERVE) 17 February 2000 (2000-02-17) see pages 16-27, examples of quantitative gene expresion of LSG, claims 1-7	1-17
Y	WO 96 02552 A (BOLLON ARTHUR P ;CYTOCLONAL PHARMACEUTICS INC (US); TORCZYNSKI RIC) 1 February 1996 (1996-02-01) the whole document	1-17
Y	WO 99 40190 A (SHIMADA YOSHIKAZU ;NAGATA MASAMI (JP); OZAKI KOUICHI (JP); HORIE M) 12 August 1999 (1999-08-12) claims 1-23 and translated description in the derived EP1074621 application.	1-17

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
6 March 2003	04.07.03

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Vix, O

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/45080

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	OZAKI K ET AL: "ISOLATION AND CHARACTERIZATION OF A NOVEL HUMAN LUNG-SPECIFIC GENE HOMOLOGOUS TO LYSOSOMAL MEMBRANE GLYCOPROTEINS 1 AND 2: SIGNIFICANTLY INCREASED EXPRESION IN CANCERS OF VARIOUS TISSUES" CANCER RESEARCH, AMERICAN ASSOCIATION FOR CANCER RESEARCH, BALTIMORE, MD, US, vol. 58, 15 August 1998 (1998-08-15), pages 3499-3503, XP002929103 ISSN: 0008-5472 the whole document	1-17

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

national application No.
PCT/US 01/45080

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claim 16 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-17 (all partially)

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1: claims 1-17 (all partially)

An isolated nucleic acid molecule comprising the nucleic acid of SEQ ID N 1, and its encoded amino acid sequence of SEQ ID N 116.

Inventions 2-115: claims 1-17 (all partially)

An isolated nucleic acid molecule comprising the nucleic acid of SEQ ID N 2 to 115, and their encoded amino acid sequence of SEQ ID N 117-208.

INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No
PCT/US 01/45080

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0008206	A	17-02-2000	CA	2347656 A1	17-02-2000
			EP	1104486 A1	06-06-2001
			JP	2002522046 T	23-07-2002
			WO	0008206 A1	17-02-2000
WO 9602552	A	01-02-1996	US	5589579 A	31-12-1996
			AU	700915 B2	14-01-1999
			AU	3359295 A	16-02-1996
			BR	9508417 A	18-11-1997
			CA	2195403 A1	01-02-1996
			EP	0804451 A1	05-11-1997
			JP	10503087 T	24-03-1998
			WO	9602552 A1	01-02-1996
			US	5773579 A	30-06-1998
WO 9940190	A	12-08-1999	CA	2319668 A1	12-08-1999
			CN	1289366 T	28-03-2001
			EP	1074621 A1	07-02-2001
			WO	9940190 A1	12-08-1999
			US	2002151025 A1	17-10-2002
			US	6403785 B1	11-06-2002